



UNICAMP

UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

ALINE CRISTINA DE SOUZA

Memórias Associativas Exponenciais Fuzzy com Aplicação em Classificação

Campinas

2018

Aline Cristina de Souza

Memórias Associativas Exponenciais Fuzzy com Aplicação em Classificação

Tese apresentada ao Instituto de Matemática,
Estatística e Computação Científica da Uni-
versidade Estadual de Campinas como parte
dos requisitos exigidos para a obtenção do
título de Doutora em Matemática Aplicada.

Orientador: Marcos Eduardo Ribeiro do Valle Mesquita

Este exemplar corresponde à versão fi-
nal da Tese defendida pela aluna Aline
Cristina de Souza e orientada pelo
Prof. Dr. Marcos Eduardo Ribeiro do
Valle Mesquita.

Campinas

2018

Agência(s) de fomento e nº(s) de processo(s): CAPES; FAPESP, 2015/00745-1

ORCID: <https://orcid.org/0000-0002-8310-2779>

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

Souza, Aline Cristina de, 1991-
So89m Memórias associativas exponenciais fuzzy com aplicação em classificação
/ Aline Cristina de Souza. – Campinas, SP : [s.n.], 2018.

Orientador: Marcos Eduardo Ribeiro do Valle Mesquita.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Memória associativa. 2. Sistemas fuzzy. 3. Classificação. 4.
Reconhecimento facial (Computação). I. Mesquita, Marcos Eduardo Ribeiro do
Valle, 1979-. II. Universidade Estadual de Campinas. Instituto de Matemática,
Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Exponential fuzzy associative memories with application in classification

Palavras-chave em inglês:

Associative memory

Fuzzy systems

Classification

Human face recognition (Computer science)

Área de concentração: Matemática Aplicada

Titulação: Doutora em Matemática Aplicada

Banca examinadora:

Marcos Eduardo Ribeiro do Valle Mesquita [Orientador]

Fernando Antonio Campos Gomide

Peter Sussner

Marley Maria Bernardes Rebuzzi Vellasco

Sandra Aparecida Sandri

Data de defesa: 11-10-2018

Programa de Pós-Graduação: Matemática Aplicada

**Tese de Doutorado defendida em 11 de outubro de 2018 e aprovada
pela banca examinadora composta pelos Profs. Drs.**

Prof(a). Dr(a). MARCOS EDUARDO RIBEIRO DO VALLE MESQUITA

Prof(a). Dr(a). FERNANDO ANTONIO CAMPOS GOMIDE

Prof(a). Dr(a). PETER SUSSNER

Prof(a). Dr(a). MARLEY MARIA BERNARDES REBUZZI VELLASCO

Prof(a). Dr(a). SANDRA APARECIDA SANDRI

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

À minha família.

Agradecimentos

Agradeço primeiramente à Deus pela minha vida, por me permitir chegar até aqui e a Nossa Senhora Aparecida por sua proteção e intercessão durante toda a caminhada.

Agradeço também toda a minha família, em especial aos meus pais Rogério e Silvia e à minha irmã Amanda, por todo amor e compreensão, por serem minha base e suporte em todos os momentos. Agradeço ao meu namorado Luiz Fernando por estar sempre ao meu lado, por me apoiar e incentivar. Vocês são graça de Deus na minha vida.

Agradeço ao Prof. Marcos Eduardo Valle, pela orientação, disponibilidade, paciência e por todos os ensinamentos compartilhados. Agradeço também à Prof.^a Marley Vellasco, à Prof.^a Sandra Sandri, ao Prof. Fernando Gomide, ao Prof. Peter Sussner e ao Prof. Estevão Esmi, pelas críticas e sugestões a este trabalho. A todos os professores que contribuíram para a minha formação, meu muito obrigada.

Agradeço a cada um dos meus amigos pela força, companheirismo e pelos bons momentos compartilhados. Em especial, agradeço aos amigos e colegas do IMECC, do Laboratório MILAB, da Pastoral Universitária da UNICAMP e do Grupo de Oração Universitário Beraká.

Agradeço também a todos os funcionários do IMECC, em especial aos funcionários da secretaria de pós-graduação e da biblioteca por todo auxílio. Agradeço ao IMECC e à UNICAMP por fornecerem toda a estrutura para a realização desta tese. A todas as pessoas que me auxiliaram direta ou indiretamente na realização deste trabalho, minha gratidão.

Por fim, gostaria de agradecer o apoio financeiro recebido durante este período. Nos primeiros meses, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Nos demais meses, o presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - Processo 2015/00745-1.

Resumo

Memórias associativas são modelos matemáticos cujo principal objetivo é armazenar e recuperar informação por associação. Tais modelos são projetados para armazenar um conjunto finito de pares, chamado conjunto das memórias fundamentais, e devem apresentar certa tolerância a ruído, isto é, serem capazes de recuperar uma certa informação armazenada mesmo a partir de uma versão incompleta ou corrompida de um item memorizado. As memórias associativas recorrentes por correlação (RCAMs, do inglês *Recurrent Correlation Associative Memories*), introduzidas por Chiueh e Goodman, apresentam grande capacidade de armazenamento e excelente tolerância a ruído. Todavia, as RCAMs são projetadas para armazenar e recuperar padrões bipolares. As memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs, do inglês *Generalized Recurrent Exponential Fuzzy Associative Memories*) podem ser vistas como uma versão generalizada das RCAMs capazes de armazenar e recuperar conjuntos *fuzzy*.

Nesta tese, introduzimos as memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs, do inglês *Generalized Exponential Bidirectional Fuzzy Associative Memories*), uma extensão das GRE-FAMs para o caso heteroassociativo. Uma vez que as GEB-FAMs são baseadas em uma medida de similaridade, realizamos um estudo de diversas medidas de similaridade da literatura, dentre elas as medidas de similaridade baseadas em cardinalidade e a medida de similaridade estrutural (SSIM). Além disso, mostramos que as GEB-FAMs exibem ótima capacidade de armazenamento e apresentamos uma caracterização da saída de um passo das GEB-FAMs quando um dos seus parâmetros tende a infinito. No entanto, em experimentos computacionais, bons resultados foram obtidos por um único passo da GEB-FAM com valores do parâmetro no intervalo $[1,10]$. Como a dinâmica das GEB-FAMs ainda não está totalmente compreendida, este fato motivou um estudo mais aprofundado das GEB-FAMs de passo único, modelos denominados memórias associativas *fuzzy* com núcleo (*fuzzy*-KAM, do inglês *fuzzy Kernel Associative Memories*). Interpretamos este modelo utilizando um núcleo *fuzzy* e propomos ajustar seu parâmetro utilizando o conceito de entropia. Apresentamos também duas abordagens para classificação de padrões usando as *fuzzy*-KAMs. Finalmente, descrevemos os experimentos computacionais realizados para avaliar o desempenho de tais abordagens em problemas de classificação e reconhecimento de faces. Na maioria dos experimentos realizados, em ambos os tipos de problemas, os classificadores definidos com base nas abordagens propostas obtiveram desempenho satisfatório e competitivo com os obtidos por outros modelos da literatura, o que mostra a versatilidade de tais abordagens.

Palavras-chave: memória associativa. sistemas *fuzzy*. medidas de similaridade. problemas de classificação. reconhecimento de faces.

Abstract

Associative memories are mathematical models whose main objective is to store and recall information by association. Such models are designed for the storage a finite set of pairs, called fundamental memory set, and they must present certain noise tolerance, that is, they should be able to retrieve a stored information even from an incomplete or corrupted version of a memorized item. The recurrent correlation associative memories (RCAMs), introduced by Chiueh and Goodman, present large storage capacity and excellent noise tolerance. However, RCAMs are designed to store and retrieve bipolar patterns. The generalized recurrent exponential fuzzy associative memories (GRE-FAMs) can be seen as a generalized version of RCAMs capable of storing and retrieving fuzzy sets.

In this thesis, we introduce the generalized exponential bidirectional fuzzy associative memories (GEB-FAMs), an extension of GRE-FAMs to the heteroassociative case. Since GEB-FAMs are based on a similarity measure, we conducted a study of several measures from the literature, including the cardinality based similarity measure and the structural similarity index (SSIM). Furthermore, we show that GEB-FAMs exhibit optimal storage capacity and we present a characterization of the output of a single-step GEB-FAM when one of its parameters tends to infinity. However, in computational experiments, good results were obtained by a single-step GEB-FAM with parameter values in the interval $[1,10]$. As the dynamics of the GEB-FAMs is still not fully understood, this fact led to a more detailed study of the single-step GEB-FAMs, referred to as fuzzy kernel associative memories (fuzzy-KAMs). We interpret this model by using a fuzzy kernel and we propose to adjust its parameter by using the concept of entropy. Also, we present two approaches to pattern classification using the fuzzy-KAMs. Finally, we describe computational experiments used to evaluate the performance of such approaches in classification and face recognition problems. In most of the experiments performed, in both types of problems, the classifiers defined based on the proposed approaches obtained satisfactory and competitive performance with those obtained by other models from the literature, which shows the versatility of such approaches.

Keywords: associative memory. fuzzy systems. similarity measures. classification problems. face recognition.

Lista de abreviaturas e siglas

AM	Memória Associativa (acrônimo do inglês <i>Associative Memory</i>)
RCAM	Memória associativa recorrente por correlação (acrônimo do inglês <i>Recurrent Correlation Associative Memory</i>)
ECAM	Memória associativa exponencial por correlação (acrônimo do inglês <i>Exponential Correlation Associative Memory</i>)
EBAM	Memória associativa exponencial bidirecional (acrônimo do inglês <i>Exponential Bidirectional Associative Memory</i>)
MERAM	Memória associativa recorrente exponencial multivalorada (acrônimo do inglês <i>Multivalued Exponential Recurrent Associative Memory</i>)
RE-FAM	Memória associativa recorrente exponencial fuzzy (acrônimo do inglês <i>Recurrent Exponential Fuzzy Associative Memory</i>)
GRE-FAM	Memória associativa recorrente exponencial fuzzy generalizada (acrônimo do inglês <i>Generalized Recurrent Exponential Fuzzy Associative Memory</i>)
GEB-FAM	Memória associativa bidirecional exponencial fuzzy generalizada (acrônimo do inglês <i>Generalized Exponential Bidirectional Fuzzy Associative Memory</i>)
SSIM	Medida de similaridade estrutural (acrônimo do inglês <i>Structural Similarity</i>)
KAM	Memória associativa com núcleo (acrônimo do inglês <i>Kernel Associative Memory</i>)
Fuzzy-KAM	Memória associativa fuzzy com núcleo (acrônimo do inglês <i>Fuzzy Kernel Associative Memory</i>)

Sumário

1	INTRODUÇÃO	11
1.1	Principais Contribuições da Tese	15
1.2	Trabalhos desenvolvidos	16
1.3	Organização da Tese	17
2	CONCEITOS PRELIMINARES	19
2.1	Conjuntos <i>fuzzy</i>	19
2.2	Medidas de Similaridade	22
3	MEMÓRIAS ASSOCIATIVAS	28
3.1	Memórias Associativas Bipolares	28
3.2	Memórias Associativas Reais	33
3.3	Memórias Associativas <i>Fuzzy</i>	37
4	MEMÓRIAS ASSOCIATIVAS BIDIRECIONAIS EXPONENCIAIS FUZZY GENERALIZADAS	42
5	MEMÓRIAS ASSOCIATIVAS FUZZY COM NÚCLEO (FUZZY- KAM)	51
5.1	Medidas de Similaridade e Núcleos <i>Fuzzy</i>	51
5.2	Definição e Propriedades de uma <i>Fuzzy-KAM</i>	52
5.3	Entropia e Ajuste do Parâmetro α	57
5.4	Formulação Matricial da <i>Fuzzy-KAM</i>	61
6	APLICAÇÕES EM PROBLEMAS DE CLASSIFICAÇÃO	68
6.1	Abordagem Baseada em Memória Autoassociativa	68
6.2	Abordagem Baseada em Memória Heteroassociativa	73
6.3	Sobre o parâmetro α	75
6.4	Experimentos Computacionais	75
6.4.1	Alguns Problemas de Classificação das Bases KEEL e UCI	76
6.4.2	Problemas de Reconhecimento de Faces	80
7	CONSIDERAÇÕES FINAIS	98
	REFERÊNCIAS	101

1 Introdução

Uma memória associativa é um sistema de entrada e saída inspirado pelo cérebro humano e projetado para armazenar e recordar um conjunto finito de pares $\{(\mathbf{a}^\xi, \mathbf{b}^\xi), \xi = 1, \dots, p\} \subset \mathcal{X} \times \mathcal{Y}$, denominado conjunto das memórias fundamentais [28]. Cada par $(\mathbf{a}^\xi, \mathbf{b}^\xi)$, para todo $\xi = 1, \dots, p$, é chamado uma memória fundamental.

Em termos matemáticos, uma memória associativa é uma aplicação $\mathcal{M} : \mathcal{X} \rightarrow \mathcal{Y}$ que deve satisfazer $\mathcal{M}(\mathbf{a}^\xi) = \mathbf{b}^\xi$ (ou $\mathcal{M}(\mathbf{a}^\xi)$ deve ser suficientemente próximo de \mathbf{b}^ξ), para todo $\xi = 1, \dots, p$. Em outras palavras, quando \mathbf{a}^ξ é apresentado como entrada, a memória deve produzir \mathbf{b}^ξ (ou um padrão suficientemente próximo de \mathbf{b}^ξ) como saída. Quando esta propriedade é satisfeita, dizemos que a memória fundamental $(\mathbf{a}^\xi, \mathbf{b}^\xi)$ foi corretamente armazenada na memória. Além disso, uma memória associativa deve apresentar certa tolerância a ruídos, isto é, ser capaz de recuperar um item armazenado \mathbf{b}^ξ mesmo quando uma versão ruidosa ou corrompida do padrão \mathbf{a}^ξ associado a ele é apresentado como entrada.

Suponhamos, por exemplo, que desejamos armazenar em uma memória associativa um conjunto de p imagens de forma que, ao apresentarmos uma dessas imagem como entrada, a memória produza a mesma imagem (ou uma imagem suficientemente próxima) como saída. Neste caso, cada imagem está associada a si própria. Portanto, denotando por \mathbf{a}^ξ , $\xi = 1, \dots, p$ as imagens que gostaríamos de armazenar, o conjunto das memórias fundamentais seria dado por $\{(\mathbf{a}^\xi, \mathbf{b}^\xi), \xi = 1, \dots, p\}$, com $\mathbf{a}^\xi = \mathbf{b}^\xi$, $\xi = 1, \dots, p$ ou, de modo equivalente, $\{(\mathbf{a}^\xi, \mathbf{a}^\xi), \xi = 1, \dots, p\}$. Neste caso, a memória é dita autoassociativa. Especificamente, dizemos que a memória é autoassociativa se $\mathbf{a}^\xi = \mathbf{b}^\xi$ para todo $\xi = 1, \dots, p$ e, neste caso, o conjunto das memórias fundamentais $\{(\mathbf{a}^\xi, \mathbf{b}^\xi), \xi = 1, \dots, p\} \subset \mathcal{X} \times \mathcal{Y}$ pode ser escrito simplesmente na forma $\{\mathbf{a}^\xi, \xi = 1, \dots, p\}$ [26, 28].

Suponhamos, agora, que desejamos armazenar em uma memória associativa um conjunto de p pares $(\mathbf{a}^\xi, \mathbf{b}^\xi)$ em que \mathbf{a}^ξ representa a imagem de uma pessoa e \mathbf{b}^ξ representa seu nome. Assim, ao apresentarmos uma imagem \mathbf{a}^ξ como entrada esperamos que a memória produza como saída o nome da pessoa a qual tal imagem pertence. Assim sendo, o conjunto das memórias fundamentais seria dado por $\{(\mathbf{a}^\xi, \mathbf{b}^\xi), \xi = 1, \dots, p\}$, com $\mathbf{a}^\xi \neq \mathbf{b}^\xi$, $\xi = 1, \dots, p$. Neste caso, a memória é dita heteroassociativa. Especificamente, dizemos que a memória é heteroassociativa se existe $\xi \in \{1, \dots, p\}$ tal que $\mathbf{a}^\xi \neq \mathbf{b}^\xi$ [26, 28].

Além disso, nos referiremos às memórias associativas projetadas para armazenamento e recordação de padrões bipolares (vetores com coordenadas -1 ou 1) como memórias associativas bipolares, às memórias associativas projetadas para armazenamento e recordação de padrões reais (vetores com coordenadas reais) como memórias associativas

reais e às memórias associativas projetadas para armazenamento e recordação de uma família finita de conjuntos *fuzzy* como memórias associativas *fuzzy* [26, 27, 74].

Memórias associativas possuem várias aplicações, dentre as quais podemos destacar os problemas de classificação e reconhecimento de padrões [19, 72, 84, 85], otimização [33], previsão [71, 73, 74] e controle [41, 43].

A rede de Hopfield [34], proposta por J.J. Hopfield em 1982, é uma rede neural recorrente com uma camada totalmente conectada que pode ser utilizada para implementar uma memória autoassociativa bipolar. Hopfield mostrou que a rede converge por meio de uma função energia, a qual é limitada inferiormente e decresce quando o estado de um neurônio é atualizado [34]. Apesar desta caracterização e das aplicações bem sucedidas, a rede de Hopfield apresenta baixa capacidade de armazenamento [51]. Motivados por esta restrição, diversos pesquisadores propuseram modelos aperfeiçoados da rede de Hopfield. Dentre eles podemos citar as memórias associativas recorrentes por correlação (RCAM, acrônimo do inglês *recurrent correlation associative memory*) [9] e, em particular, as memórias associativas exponenciais por correlação (ECAM, acrônimo do inglês *exponential correlation associative memory*) [9], modelos autoassociativos propostos por Chiueh e Goodman em 1991. Para o caso heteroassociativo, Kosko apresentou em 1988 a memória associativa bidirecional (BAM, acrônimo do inglês *bidirectional associative memory*) [42] como uma generalização da rede de Hopfield. Em 1993, Jeng et al propuseram a memória associativa bidirecional exponencial (EBAM, acrônimo do inglês *exponential bidirectional associative memory*) [37], modelo que pode ser visto como uma generalização da ECAM e da BAM.

Assim como a rede de Hopfield, tanto a ECAM quanto a EBAM são modelos projetados para o armazenamento e recordação de vetores bipolares. Todavia, em algumas aplicações, o armazenamento de vetores multivalorados ou conjuntos *fuzzy* é necessário. Este fato motivou o surgimento de modelos de memórias associativas para armazenamento de vetores reais, dentre os quais podemos destacar as memórias associativas recorrentes exponenciais multivaloradas [10] (MERAMs, acrônimo do inglês *multivalued exponential recurrent associative memory*) e as memórias associativas com núcleo [84] (KAMs, acrônimo do inglês *kernel associative memories*), modelos que, num certo sentido, podem ser vistos como uma extensão das RCAMs para o armazenamento e recordação de vetores reais.

Em 2014, Valle introduziu as memórias associativas recorrentes exponenciais *fuzzy* (RE-FAMs, acrônimo do inglês *recurrent exponential fuzzy associative memories*) [75], modelo autoassociativo que pode ser visto com uma versão *fuzzy* da ECAM. No entanto, devido à interferência cruzada entre os padrões a serem armazenados, as RE-FAMs nem sempre armazenam corretamente todas as memórias fundamentais e nestes casos, não definem uma memória associativa. Diante disso, as memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs, acrônimo do inglês *generalized recurrent*

exponential fuzzy associative memories) [64, 70] foram desenvolvidas. Brevemente, as GRE-FAMs são obtidas das RE-FAMs por meio do acréscimo de uma camada de neurônios lineares com o objetivo de reduzir a interferência cruzada entre as memórias fundamentais. Outros modelos de memórias associativas *fuzzy* que apresentam certas relações com as GRE-FAMs são as θ -FAMs (acrônimo do inglês *θ -fuzzy associative memories* [17] e as TE-FAMs [18] (acrônimo do inglês *tunable equivalence fuzzy associative memories*).

A Figura 1 apresenta um diagrama que resume a trajetória de estudo dos principais modelos que motivaram o desenvolvimento, no contexto deste trabalho, das memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs, acrônimo do inglês *generalized exponential bidirectional fuzzy associative memories*) e das memórias associativas *fuzzy* com núcleo (*fuzzy*-KAMs, acrônimo do inglês *fuzzy kernel associative memories*), modelos que estão entre as principais contribuições desta tese. No diagrama, os modelos bipolares estão destacados em azul e os modelos reais e *fuzzy* em vermelho. Além disso, os modelos recorrentes nas caixas com borda contínua, modelos recorrentes bidirecionais nas caixas com bordas tracejadas e nas caixas com bordas pontilhadas os modelos não-recorrentes.

Assim como no caso das GRE-FAMs, a definição das GEB-FAMs exige uma medida de similaridade (*fuzzy*). De modo geral, uma medida de similaridade é uma função que associa a cada par de conjuntos *fuzzy* um número que representa o grau de igualdade entre esses conjuntos. No entanto, a definição de medida de similaridade não é única. Uma definição abrangente foi proposta por De Baets e De Meyer em [12], a qual engloba as medidas de similaridade baseadas em cardinalidade e uma versão normalizada da medida de similaridade proposta por Xuecheng [81]. No mesmo período, Wang et al [79] propuseram uma medida de similaridade estrutural, conhecida por SSIM (do inglês *structural similarity*). A medida estrutural SSIM, desenvolvida principalmente para avaliar a similaridade entre imagens, é caracterizada por considerar a informação estrutural da imagem, independentemente de iluminação e contraste [79].

Além da medida de similaridade, as GEB-FAMs requerem em sua definição um parâmetro real positivo α . Tal parâmetro tem um papel importante na capacidade de generalização e convergência da GEB-FAM. Obtivemos uma caracterização para o primeiro passo de uma GEB-FAM quando α tende a infinito, no entanto, os melhores resultados nos experimentos computacionais foram obtidos para valores pequenos deste parâmetro (em geral, entre 1 e 10). Além disso, por meio de um exemplo, concluímos que a sequência produzida pela GEB-FAM pode não ser convergente se α não for suficientemente grande. Este fato motivou um estudo aprofundado do primeiro passo da GEB-FAM. Tal estudo foi feito utilizando uma interpretação baseada em núcleos *fuzzy* e o modelo obtido é denominado *fuzzy*-KAM.

O uso de núcleos em métodos de aprendizagem e estimação tem se tornado

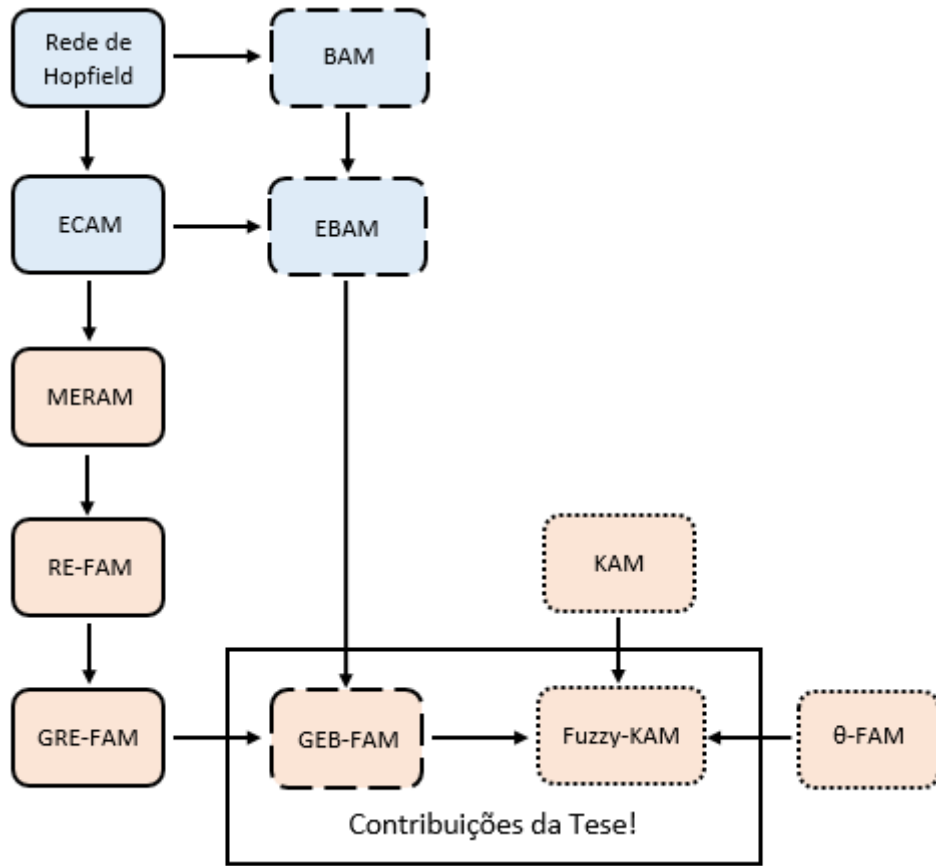


Figura 1 – Principais modelos que motivaram o desenvolvimento da tese. Modelos bipolares estão destacados em azul e modelos reais e *fuzzy* em vermelho. Modelos recorrentes nas caixas com borda contínua, modelos recorrentes bidirecionais nas caixas com bordas tracejadas e modelos não-recorrentes nas caixas com bordas pontilhadas.

cada vez mais frequente [22,31,86]. Isto se deve ao fato de que um núcleo (positivo definido) pode ser interpretado como um produto interno em espaço de alta dimensão, chamado espaço de característica. Esta interpretação é útil pois possibilita que problemas não lineares (no espaço original) sejam abordados por meio de técnicas lineares em um espaço de características de dimensão suficientemente grande, mas de forma que os cálculos sejam efetuados com os elementos do espaço original [31]. Informalmente, um núcleo pode ser definido como uma medida de similaridade. Interpretaremos uma medida de similaridade *fuzzy* como um núcleo *fuzzy* κ . O núcleo *fuzzy* κ depende do parâmetro α . Precisamente, o parâmetro α controla a capacidade do núcleo *fuzzy* de distinguir entre dois conjuntos *fuzzy* [67]. Com o objetivo de maximizar a capacidade do núcleo *fuzzy* κ em discriminar dois estímulos diferentes, propomos uma estratégia para o ajuste de α utilizando o conceito de entropia [62,67].

Após o estudo de características teóricas, investigamos a aplicação das *fuzzy*-KAMs em problemas de classificação e reconhecimento de faces. Problemas de classificação

são frequentes em nosso dia a dia. Na área da saúde, por exemplo, um tumor pode ser classificado em benigno ou maligno ou um paciente pode ser classificado como portador ou não de uma determinada doença de acordo com seus sintomas. Em outras áreas, podemos citar o exemplo de bancos que classificam um cliente como aptos ou não para receber um empréstimo de acordo com seu histórico de crédito. Outros exemplos incluem problemas de classificação de plantas, reconhecimento de padrões e reconhecimento de faces. Propusemos duas abordagens para classificação de padrões utilizando *fuzzy*-KAMs. A primeira se trata de uma abordagem autoassociativa inspirada pelos classificadores de representação esparsa [80]. Tais classificadores são desenvolvidos sob a hipótese de que uma amostra pertencente a uma determinada classe pode ser escrita aproximadamente como combinação linear dos padrões de treinamento desta classe [80]. A segunda abordagem é definida utilizando a versão heteroassociativa da *fuzzy*-KAM.

1.1 Principais Contribuições da Tese

Nesta tese, introduzimos a definição das memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs), extensão das GRE-FAMs para o caso heteroassociativo, e apresentamos alguns resultados teóricos sobre estes modelos. Especificamente, mostramos que, sob certas condições, as GEB-FAMs armazenam corretamente todas as memórias fundamentais e apresentamos uma caracterização para a saída de um passo da GEB-FAM quando um dos parâmetros do modelo tende a infinito.

Introduzimos também as memórias associativas *fuzzy* com núcleo (*fuzzy*-KAM) e apresentamos reformulações dos resultados teóricos das GEB-FAMs para as *fuzzy*-KAMs. Além disso, definimos uma condição suficiente para que a matriz K presente na definição da *fuzzy*-KAM seja definida positiva. Propusemos o ajuste do parâmetro α da *fuzzy*-KAM por meio do conceito de entropia, apresentamos uma formulação matricial e uma sugestão para definir a matriz G , necessária para a definição da *fuzzy*-KAM, com estrutura esparsa em blocos.

Fizemos um estudo de várias medidas de similaridade da literatura, dentre elas as medidas de similaridade baseadas em cardinalidade [12], uma versão normalizada da medida de similaridade proposta por Xuecheng [81] e a medida de similaridade estrutural [79].

Apresentamos a aplicação das *fuzzy*-KAMs em problemas de classificação e reconhecimento de faces. Especificamente, propusemos duas abordagens para classificação baseada nas *fuzzy*-KAMs. A primeira, baseada na versão autoassociativa e a segunda na versão heteroassociativa do modelo.

1.2 Trabalhos desenvolvidos

Como frutos do trabalho realizado para o desenvolvimento desta tese, durante o período do doutorado, foram produzidos um capítulo de livro [76], um artigo de revista [68] e quatro artigos apresentados em congressos [65–67, 69]. Estes trabalhos contém as aplicações dos modelos propostos (GEB-FAM e *fuzzy*-KAM) e das GRE-FAMs em problemas de classificação, conforme as abordagens que serão descritas no Capítulo 6, bem como alguns dos testes (ou variações destes) que descreveremos na Seção 6.4.

Especificamente, no capítulo “*Pattern Classification using Generalized Recurrent Exponential Fuzzy Associative Memories*” apresentamos um classificador (baseado na GRE-FAM) definido por meio da abordagem autoassociativa, descrita na seção 6.1, e os resultados dos testes realizados com quinze problemas de classificação.

O trabalho “Memórias Associativas Recorrentes Exponenciais *Fuzzy* Generalizadas Aplicadas à Classificação de Padrões” [66], apresentado em 2016 no XXXVI Congresso Nacional de Matemática Aplicada e Computacional (XXXVI CNMAC), também contém uma aplicação das GRE-FAMs a problemas de classificação. Neste trabalho, consideramos seis conjuntos de dados.

Em “Memória Associativa Bidirecional Exponencial *Fuzzy* Generalizada Aplicada ao Reconhecimento de Faces.” [65], apresentado em 2016 no IV Congresso Brasileiro de Sistemas *Fuzzy* (IV CBSF), apresentamos a aplicação das GEB-FAMs a um problema de reconhecimento de faces, considerando um classificador definido com base na abordagem heteroassociativa, apresentado na Seção 6.2.

No artigo “*Generalized Exponential Bidirectional Fuzzy Associative Memory with Fuzzy Cardinality-Based Similarity Measures Applied to Face Recognition*” [68], publicado na revista TEMA, consideramos as medidas de similaridade baseadas em cardinalidade e avaliamos o desempenho do classificador heteroassociativo com essas medidas em um problema de reconhecimento facial.

No trabalho “Memória Associativa Bidirecional Exponencial *Fuzzy* Generalizada com Medida de Similaridade Estrutural Aplicada a um Problema de Reconhecimento de Faces”, apresentado em 2017 no XXXVI Congresso Nacional de Matemática Aplicada e Computacional (XXXVI CNMAC), [69] consideramos o classificador heteroassociativo com a medida de similaridade estrutural SSIM.

Por fim, em “*Fuzzy kernel associative memories with application in classification*” [67], apresentado em 2018 no *North American Fuzzy Information Processing Society Meeting* (NAFIPS), apresentamos a aplicação das *fuzzy*-KAMs a problemas de classificação.

1.3 Organização da Tese

Esta tese está organizada da seguinte forma. O Capítulo 2 apresenta uma breve revisão de conceitos essenciais para o desenvolvimento desta tese. Especificamente, iniciaremos pela recordação de alguns conceitos relacionados a conjuntos *fuzzy*. Em seguida, faremos uma revisão sobre medidas de similaridade, a qual contemplará várias definições propostas na literatura, dentre elas a definição proposta por De Baets e De Meyer em [12], as medidas de similaridade baseadas em cardinalidade [12], a medida de similaridade estrutural (SSIM) [79] e uma versão normalizada da definição de medida de similaridade proposta por Xuecheng [81].

Posteriormente, no Capítulo 3, faremos uma revisão bibliográfica contendo os principais modelos da literatura que motivaram o desenvolvimento deste trabalho. Primeiramente, apresentaremos memórias associativas bipolares, a saber: a rede de Hopfield [34], as memórias associativas recorrentes por correlação (RCAMs) [9], as memórias associativas exponenciais por correlação [9] e as memórias associativas bidirecionais exponenciais (EBAMs) [37]. Em seguida, as memórias associativas recorrentes exponenciais multivariadas (MERAMs) [10] e as memórias associativas com núcleo (KAMs) [84], modelos para armazenamento e recordação de vetores com componentes reais. Finalizaremos este capítulo revisando alguns modelos de memórias associativas *fuzzy*, dentre eles, as memórias associativas recorrentes exponenciais *fuzzy* (RE-FAMs) [75], as memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs) [64, 70], as Θ -FAMs [17] e as TE-FAMs [18].

No Capítulo 4, introduziremos as memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs), modelo que estende as GRE-FAMs para o caso heteroassociativo, e apresentaremos alguns resultados teóricos sobre este modelo. Mostraremos que as GEB-FAMs, sob certas condições, possuem capacidade ótima de armazenamento e apresentaremos uma caracterização da saída da GEB-FAM quando o parâmetro α do modelo tende a infinito.

No Capítulo 5, focaremos no estudo das memórias associativas *fuzzy* com núcleo (*fuzzy*-KAM), modelos que derivam de uma GEB-FAM de passo único por meio de uma interpretação utilizando núcleos *fuzzy*. Determinaremos um limitante inferior para o parâmetro α da *fuzzy*-KAM, a partir do qual a matriz K , utilizada para a definição da *fuzzy*-KAM, é definida positiva. Posteriormente, proporemos uma estratégia para o ajuste do parâmetro α com base no conceito de entropia. Finalmente, apresentaremos uma formulação matricial para a *fuzzy*-KAM e sugerimos uma forma alternativa de definir a matriz G (presente na definição da *fuzzy*-KAM), a qual pode ser útil sobretudo nos problemas com um grande número de memórias fundamentais.

As aplicações em problemas de classificação e reconhecimento de faces serão

apresentadas no Capítulo 6. Proporemos duas abordagens para classificação de padrões utilizando *fuzzy*-KAMs. Além disso, descreveremos os experimentos realizados para avaliar o desempenho dos classificadores baseados nas *fuzzy*-KAMs, bem como os resultados obtidos. Os testes em problemas de classificação foram realizados com conjuntos de dados disponíveis nos repositórios KEEL [4] e UCI [6]. Os experimentos em reconhecimento de faces foram desenvolvidos com três bases de imagens faciais, a saber: a base de imagens faciais AR [49], a base de imagens ORL [60] e a base GT [1].

Concluiremos esta tese com as considerações finais no Capítulo 7.

2 Conceitos Preliminares

Neste capítulo, apresentaremos conceitos importantes para o desenvolvimento dos capítulos posteriores. Primeiramente, recordaremos as definições de conjuntos *fuzzy*, relação *fuzzy*, normas triangulares (t-normas), dentre outras. Posteriormente, apresentaremos algumas definições e exemplos de medidas de similaridade da literatura.

2.1 Conjuntos *fuzzy*

Todo conjunto (clássico) X pode ser definido por meio de sua função característica $\chi : U \rightarrow \{0, 1\}$, a qual indica se um elemento do conjunto universo U pertence ou não a X . A função característica é definida por:

$$\chi(x) = \begin{cases} 1, & \text{se } x \in X \\ 0, & \text{se } x \notin X. \end{cases} \quad (2.1)$$

Por exemplo, consideremos o conjunto dos números reais maiores do que 1, $X = \{x \in \mathbb{R} : x > 1\}$. Sabemos que $2 \in X$ enquanto que $0,5 \notin X$. Portanto $\chi(2) = 1$ e $\chi(0,5) = 0$.

No caso dos conjuntos clássicos, dado um elemento $x \in U$, existem apenas duas possibilidades: x pertence ao conjunto X ou não. No entanto, se considerarmos o conjunto $\tilde{X} = \{x \in \mathbb{R} : x \text{ é próximo de } 1\}$ não é tão claro quais elementos pertencem a este conjunto. No entanto, podemos afirmar, por exemplo, que 0,5 é mais próximo de 1 do que 2. Em outras palavras, podemos dizer que o grau de pertinência de 0,5 a \tilde{X} é maior que o grau de pertinência de 2 a este conjunto.

A teoria dos conjuntos *fuzzy*, proposta por Lotfi A. Zadeh em 1965 [83], possibilitou que termos imprecisos como “aproximadamente” e “em torno de” fossem tratados matematicamente. Em termos gerais, um conjunto *fuzzy* é caracterizado por meio de uma função $X : U \rightarrow [0, 1]$, chamada função de pertinência. Especificamente, um conjunto *fuzzy* é definido da seguinte forma [8, 83]:

Definição 1. *Consideremos um conjunto universo de discurso U . Um conjunto fuzzy X (subconjunto fuzzy de U) é definido por meio de uma função $X : U \rightarrow [0, 1]$, em que $X(u)$ é o grau de pertinência do elemento $u \in U$ ao conjunto fuzzy X .*

A função de pertinência associa a cada elemento $u \in U$ seu grau de pertinência ao subconjunto fuzzy X . No exemplo anterior, em que consideramos $\tilde{X} = \{x \in \mathbb{R} : x \text{ é próximo de } 1\}$, teríamos $\tilde{X}(0,5) > \tilde{X}(2)$.

Nesta tese, consideraremos apenas subconjuntos *fuzzy* de um universo de discurso finito, digamos $U = \{u_1, \dots, u_n\}$. Neste caso, podemos identificar um subconjunto *fuzzy* X de U com um vetor $X = [X(u_1), \dots, X(u_n)]^T$ pertencente ao hipercubo unitário n -dimensional $[0, 1]^n$ [43]. Além disso, cada conjunto *fuzzy* está associado a uma única função de pertinência e cada função de pertinência representa um único conjunto *fuzzy*. Por este motivo e por questões de simplicidade, utilizaremos a mesma notação, X , para nos referir ao conjunto *fuzzy* X , à sua função de pertinência $X : U \rightarrow [0, 1]$ e ao vetor $X = [X(u_1), \dots, X(u_n)]^T$. Denotaremos por $\mathcal{F}(U)$ o conjunto dos subconjuntos *fuzzy* de U e por $\mathcal{P}(U)$ o conjunto das partes de U .

A seguir, recordaremos alguns conceitos relacionados a conjuntos *fuzzy*. Dentre eles, as definições de cardinalidade e complemento de um conjunto *fuzzy*. As definições de normas triangulares e das operações de interseção, união, diferença, diferença simétrica e a definição de relação binária *fuzzy* também serão apresentadas.

Definição 2 (Cardinalidade de um conjunto *fuzzy*). *Seja X um subconjunto fuzzy de $U = \{u_1, \dots, u_n\}$. A cardinalidade de X é definida por:*

$$\text{Card}(X) = \sum_{i=1}^n X(u_i). \quad (2.2)$$

Definição 3 (Complemento de um conjunto *fuzzy*). *Seja X um subconjunto fuzzy de U . O complementar de X , denotado por X^c , é o subconjunto fuzzy de U definido pela função de pertinência $X^c : U \rightarrow [0, 1]$ dada por:*

$$X^c(u) = 1 - X(u), \text{ para todo } u \in U. \quad (2.3)$$

A noção de complemento pode ser generalizada por meio de uma negação *fuzzy* [57].

Definição 4 (T-normas). *Uma norma triangular ou, simplesmente t-norma, é uma função $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ que satisfaz as seguintes propriedades, para quaisquer $x, y, z, w \in [0, 1]$ [38]:*

$$T_1 \text{ (Identidade): } T(x, 1) = x;$$

$$T_2 \text{ (Comutatividade): } T(x, y) = T(y, x);$$

$$T_3 \text{ (Associatividade): } T(x, T(y, z)) = T(T(x, y), z);$$

$$T_4 \text{ (Monotonicidade): Se } x \leq z \text{ e } y \leq w \text{ então } T(x, y) \leq T(z, w);$$

Exemplo 1. *Alguns exemplos de t-norma são dados por [38]:*

- *t-norma do Mínimo: $M : [0, 1] \times [0, 1] \rightarrow [0, 1]$ definida por $M(x, y) = \min\{x, y\}$.*

- *t-norma do Produto*: $P : [0, 1] \times [0, 1] \rightarrow [0, 1]$ definida por $P(x, y) = xy$.
- *t-norma de Łukasiewicz*: $W : [0, 1] \times [0, 1] \rightarrow [0, 1]$ definida por $W(x, y) = \max\{x + y - 1, 0\}$.
- *t-norma drástica*: $Z : [0, 1] \times [0, 1] \rightarrow [0, 1]$ definida por

$$Z(x, y) = \begin{cases} \min\{x, y\}, & \text{se } x = 1 \text{ ou } y = 1, \\ 0, & \text{caso contrário.} \end{cases}$$

- *t-norma de Frank* [21, 55]: $T_s : [0, 1] \times [0, 1] \rightarrow [0, 1]$, $s \in [0, \infty]$, é definida pela equação:

$$T_s(x, y) = \log_s \left[1 + \frac{(s^x - 1)(s^y - 1)}{s - 1} \right], \quad s > 0, s \neq 1.$$

com as seguintes identidades [12]:

- (i) $T_0(x, y) = \lim_{s \rightarrow 0} T_s(x, y) = M(x, y) = \min\{x, y\}$,
- (ii) $T_1(x, y) = \lim_{s \rightarrow 1} T_s(x, y) = P(x, y) = xy$,
- (iii) $T_\infty(x, y) = \lim_{s \rightarrow \infty} T_s(x, y) = W(x, y) = \max\{x + y - 1, 0\}$.

Dados A e B subconjuntos *fuzzy* de um conjunto universo U , as definições a seguir mostram como a interseção, a união, a diferença e a diferença simétrica de A e B podem ser calculadas por meio de t-normas.

Definição 5 (Interseção de conjuntos *fuzzy*). *Sejam A e B subconjuntos fuzzy de U e $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ uma t-norma. A interseção de A e B é definida por:*

$$(A \cap B)(u) = T(A(u), B(u)), \quad \forall u \in U. \quad (2.4)$$

No caso mais geral, a interseção de conjuntos *fuzzy* pode ser definida usando uma conjunção *fuzzy* [7, 8, 57].

Definição 6 (União de conjuntos *fuzzy*). *Sejam A e B subconjuntos fuzzy de U e $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ uma t-norma de Frank. A união de A e B é definida por:*

$$(A \cup B)(u) = A(u) + B(u) - T(A(u), B(u)), \quad \forall u \in U. \quad (2.5)$$

Esta definição de união de conjuntos *fuzzy* com base em uma t-norma de Frank é considerada em [12] pois, nestas condições, a igualdade $\text{Card}(A \cup B) + \text{Card}(A \cap B) = \text{Card}(A) + \text{Card}(B)$ é válida. No entanto, no caso geral, a união de conjuntos *fuzzy* pode ser definida usando uma t-conorma, ou ainda, usando uma disjunção *fuzzy* [7, 8, 57].

Definição 7 (Diferença de conjuntos *fuzzy*). Sejam A e B subconjuntos *fuzzy* de U e $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ uma t -norma. A diferença de A e B é definida por:

$$(A \setminus B)(u) = A(u) - T(A(u), B(u)), \quad \forall u \in U. \quad (2.6)$$

Definição 8 (Diferença simétrica de conjuntos *fuzzy*). Sejam A e B subconjuntos *fuzzy* de U e $T : [0, 1] \times [0, 1] \rightarrow [0, 1]$ uma t -norma. A diferença simétrica de A e B é definida por:

$$(A \Delta B)(u) = A(u) + B(u) - 2T(A(u), B(u)), \quad \forall u \in U. \quad (2.7)$$

A seguir apresentaremos a definição de relação binária *fuzzy*.

Definição 9 (Relação binária *fuzzy*). Seja U um conjunto clássico. Uma relação binária *fuzzy* sobre U é caracterizada por uma função $R : U \times U \rightarrow [0, 1]$.

Observação 1. Note que uma relação *fuzzy* $R : U \times U \rightarrow [0, 1]$ pode ser vista como um subconjunto *fuzzy* de $U \times U$.

Definição 10. Seja T uma t -norma. Uma relação binária *fuzzy* R é dita **T-transitiva** se satisfaz a seguinte equação para quaisquer $x, y, z \in U$:

$$T(R(x, y), R(y, z)) \leq R(x, z).$$

Além disso, dizemos que uma relação binária *fuzzy* $R : U \times U \rightarrow [0, 1]$ é:

- **simétrica** se $R(x, y) = R(y, x)$, para quaisquer $x, y \in U$;
- **reflexiva** se $R(x, x) = 1$, para qualquer $x \in U$;
- **localmente reflexiva** se $R(x, x) \geq R(x, y)$ para quaisquer $x, y \in U$ [12].

Na próxima seção faremos uma revisão sobre várias definições de medidas de similaridade presentes na literatura.

2.2 Medidas de Similaridade

Uma medida de similaridade (*fuzzy*), de modo geral, é uma função que associa a cada par de conjuntos *fuzzy* um número que indica o quanto esses conjuntos são iguais. Existem na literatura várias definições de medida de similaridade [12, 81]. Nesta seção apresentaremos algumas dessas definições, as quais foram utilizadas nos experimentos computacionais descritos no Capítulo 6.

Uma definição abrangente de medida de similaridade foi proposta por De Baets e De Meyer em [12]:

Definição 11. Uma medida de similaridade fuzzy é uma relação binária fuzzy simétrica sobre $\mathcal{F}(U)$, isto é, uma função $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ tal que $\mathcal{S}(A, B) = \mathcal{S}(B, A)$, para quaisquer conjuntos fuzzy A e B pertencentes a $\mathcal{F}(U)$.

Uma medida de similaridade fuzzy $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ é dita reflexiva se a propriedade $\mathcal{S}(A, A) = 1$ é válida para qualquer conjunto fuzzy $A \in \mathcal{F}(U)$ e é dita localmente reflexiva se $\mathcal{S}(A, A) \geq \mathcal{S}(A, B)$ para quaisquer $A, B \in \mathcal{F}(U)$ [12]. Além disso, dizemos que uma medida de similaridade é forte quando $\mathcal{S}(A, B) = 1$ se, e somente se, $A = B$. Por fim, dada uma t -norma T , dizemos que uma medida de similaridade fuzzy é T -transitiva [12] se satisfaz a Definição 10, isto é, se satisfaz a seguinte equação para quaisquer conjuntos fuzzy $A, B, C \in \mathcal{F}(U)$:

$$T(\mathcal{S}(A, B), \mathcal{S}(B, C)) \leq \mathcal{S}(A, C).$$

A definição proposta por De Baets e De Meyer em [12], engloba outras definições da literatura como por exemplo as medidas de similaridade baseadas em cardinalidade [12], e a medida de similaridade proposta por Xuecheng [81], as quais serão apresentadas a seguir.

De Baets e De Meyer propuseram em [12] uma classe de medidas de similaridade fuzzy racionais baseadas na cardinalidade dos conjuntos fuzzy envolvidos. Tais medidas são obtidas por meio da fuzzificação das medidas de similaridade racionais baseadas em cardinalidade proposta para conjuntos clássicos em [13]. Precisamente, considerando um universo de discurso finito U , dados $A, B \in \mathcal{F}(U)$, tais medidas são definidas por [12]:

$$\mathcal{S}(A, B) = \frac{a\alpha_{A,B} + b\omega_{A,B} + c\delta_{A,B} + d\nu_{A,B}}{a'\alpha_{A,B} + b'\omega_{A,B} + c'\delta_{A,B} + d'\nu_{A,B}}. \quad (2.8)$$

em que $a, a', b, b', c, c', d, d' \in \{0, 1\}$ são parâmetros,

- (i) $\alpha_{A,B} = \min\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\},$
- (ii) $\omega_{A,B} = \max\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\},$
- (iii) $\delta_{A,B} = \text{Card}(A \cap B),$
- (iv) $\nu_{A,B} = \text{Card}(A \cup B)^c.$

e tais que a interseção, união, diferença e diferença simétrica dos conjuntos fuzzy são calculados, respectivamente, por meio das equações (2.4), (2.5), (2.6) e (2.7), baseadas em uma t -norma de Frank.

A Tabela 1, extraída de [12], apresenta diversas medidas obtidas por meio da expressão (2.8) e dos esquemas de fuzzificação dados pelas Equações (2.4), (2.5), (2.6), (2.7), baseados em t -normas de Frank. Especificamente, nos concentramos nas medidas

Tabela 1 – Expressões das medidas de similaridade racionais baseadas em cardinalidade.
Tabela extraída de [12].

\mathcal{S}	Expressão	\mathcal{S}	Expressão
R_1	$\frac{\text{Card}(A \cap B)}{\max\{\text{Card}(A), \text{Card}(B)\}}$	R_{10}	$\frac{\min\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\}}{\max\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\}}$
R_2	$\frac{\text{Card}(A \Delta B)^c}{\max\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}$	R_{11}	$\frac{\min\{\text{Card}(A), \text{Card}(B)\}}{\max\{\text{Card}(A), \text{Card}(B)\}}$
R_3	$\frac{\text{Card}(A \cap B)}{\min\{\text{Card}(A), \text{Card}(B)\}}$	R_{12}	$\frac{\min\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}{\max\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}$
R_4	$\frac{\text{Card}(A \Delta B)^c}{\min\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}$	R_{13}	$\frac{\min\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\}}{\text{Card}(A \Delta B)}$
R_5	$\frac{\text{Card}(A \cap B)}{\text{Card}(A \cup B)}$	R_{14}	$\frac{\min\{\text{Card}(A), \text{Card}(B)\}}{\text{Card}(A \cup B)}$
R_6	$\frac{\text{Card}(A \Delta B)^c}{n}$	R_{15}	$\frac{\min\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}{n}$
R_7	$\frac{\max\{\text{Card}(A \setminus B), \text{Card}(B \setminus A)\}}{\text{Card}(A \Delta B)}$	S_{17}	$\frac{\text{Card}(A \cap B)}{\max\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}$
R_8	$\frac{\max\{\text{Card}(A), \text{Card}(B)\}}{\text{Card}(A \cup B)}$	S_{18}	$\frac{\text{Card}(A \cap B)}{n}$
R_9	$\frac{\max\{\text{Card}(A \setminus B)^c, \text{Card}(B \setminus A)^c\}}{n}$	S_{19}	$\frac{\min\{\text{Card}(A), \text{Card}(B)\}}{n}$

de similaridade mostradas na Tabela 1 obtidas considerando as t -normas de Frank dadas pelo mínimo $T_0(x, y) = M(x, y) = \min(x, y)$, pelo produto $T_1(x, y) = P(x, y) = xy$ e pela t -norma de Łukasiewicz $T_\infty(x, y) = W(x, y) = \max\{x + y - 1, 0\}$.

Observação 2. As medidas de similaridade apresentadas na Tabela 1 satisfazem as propriedades [12]:

- (i) $R_1, R_5, R_6, R_{13}, R_{15}$ e S_{18} são W -transitivas,
- (ii) R_{11} e R_{14} são P -transitivas,
- (iii) S_{19} é M -transitiva,
- (iv) R_2 é Z -transitiva,
- (v) R_i , para $i \in \{1, 2, 5, 6, 13, 14, 15\}$, são reflexivas quando consideramos $T = M$,

- (vi) R_{11} é reflexiva,
- (vii) S_{18} é localmente reflexiva quando consideramos $T = M$,
- (viii) S_{19} é localmente reflexiva.

Outra definição de medida de similaridade foi proposta por Xuecheng em [81]. Uma versão normalizada desta definição é apresentada a seguir:

Definição 12 (Medida de similaridade de Xuecheng). *Uma medida de similaridade de Xuecheng é uma função $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ que satisfaz, para quaisquer conjuntos fuzzy $A, B, C, D \in \mathcal{F}(U)$, as seguintes propriedades:*

1. $\mathcal{S}(A, B) = \mathcal{S}(B, A)$.
2. $\mathcal{S}(A, A) = 1$.
3. Se $A \subseteq B \subseteq C \subseteq D$, então $\mathcal{S}(A, D) \leq \mathcal{S}(B, C)$.
4. $\mathcal{S}(A, \bar{A}) = 0$, para todo subconjunto clássico $A \in \mathcal{P}(U)$.

Note que uma medida de similaridade de Xuecheng é uma medida de similaridade reflexiva.

A seguir, alguns exemplos de medidas de similaridade que satisfazem a definição de Xuecheng.

Exemplo 2. *Consideremos um conjunto universo finito $U = \{u_1, \dots, u_n\}$. As funções definidas abaixo são exemplos de medidas de similaridade forte [16, 88]:*

1. *Medida de Similaridade de Gregson (também chamada de medida de similaridade de Jaccard) [36]*

$$\mathcal{S}_G(A, B) = \begin{cases} \frac{\sum_{j=1}^n A(u_j) \wedge B(u_j)}{\sum_{j=1}^n A(u_j) \vee B(u_j)}, & \text{se } A \cup B \neq \emptyset \\ 1, & \text{se } A \cup B = \emptyset, \end{cases} \quad (2.9)$$

onde os símbolos “ \wedge ” and “ \vee ” denotam respectivamente as operações mínimo e máximo.

2. Medida de Similaridade de Eisler e Ekman

$$\mathcal{S}_E(A, B) = \begin{cases} \frac{2 \sum_{j=1}^n A(u_j) \wedge B(u_j)}{\sum_{j=1}^n A(u_j) + \sum_{j=1}^n B(u_j)}, & \text{se } A \cup B \neq \emptyset \\ 1, & \text{se } A \cup B = \emptyset, \end{cases} \quad (2.10)$$

3. Medida de Similaridade baseada na norma-1

$$\mathcal{S}_H(A, B) = 1 - \frac{1}{n} \sum_{j=1}^n |A(u_j) - B(u_j)|. \quad (2.11)$$

Observação 3. Note que a medida de similaridade de Gregson dada por (2.9), também satisfaz a definição das medidas de similaridade baseadas em cardinalidade [12] descritas anteriormente.

Uma medida de similaridade desenvolvida principalmente para avaliar a similaridade entre imagens foi proposta em [79]. Tal medida, denominada *structural similarity index* (SSIM), baseia-se na informação estrutural da imagem, isto é, nos atributos relacionados à estrutura dos objetos independentemente da iluminação média e do contraste [79].

Especificamente, a medida de similaridade estrutural (SSIM) é definida a partir de uma medida de similaridade estrutural local, denotada por S_l [79]. De forma geral, S_l é calculada por meio de comparações da iluminação, do contraste e da estrutura das imagens em uma determinada janela das imagens. Em termos matemáticos, sejam $\mathbf{x} = \{x_i : i = 1, \dots, N\}$ e $\mathbf{y} = \{y_i : i = 1, \dots, N\}$ fragmentos extraídos do mesmo local de duas imagens a serem comparadas. A medida de similaridade estrutural local é definida pela expressão [79]:

$$S_l(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (2.12)$$

onde C_1 e C_2 são constantes positivas,

$$\mu_x = \frac{1}{N} \sum_{i=1}^N \omega_i x_i, \quad \sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N \omega_i (x_i - \mu_x)^2 \right)^{\frac{1}{2}}, \quad \sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N \omega_i (x_i - \mu_x)(y_i - \mu_y) \quad (2.13)$$

e $\Omega = \{\omega_i : i = 1, \dots, N\}$ é uma função de ponderação Gaussiana circular simétrica 11×11 , normalizada de modo que $\sum_{i=1}^N \omega_i = 1$ e com desvio padrão 1,5, conforme sugerido em [79].

A medida de similaridade estrutural SSIM de duas imagens X e Y é então definida por:

$$S(X, Y) = \frac{1}{M} \sum_{j=1}^M S_l(\mathbf{x}_j, \mathbf{y}_j), \quad (2.14)$$

em que X e Y são as imagens a serem comparadas, \mathbf{x}_j e \mathbf{y}_j são os fragmentos das imagens contidos na j -ésima janela local e M é o número de janelas locais das imagens. Em outras palavras, a medida de similaridade estrutural SSIM é dada pela média da medida estrutural local de todos os fragmentos.

Observemos que a medida SSIM é simétrica, limitada e reflexiva, isto é, $S(X, Y) = S(Y, X)$, $-1 \leq S(X, Y) \leq 1$ e $S(X, Y) = 1 \iff X = Y$ [78, 79]. Note que \mathcal{S} não é uma medida de similaridade *fuzzy*. Porém, as funções $\sigma(X, Y) = (S(X, Y) + 1)/2$ e $\kappa(X, Y) = e^{\alpha(S(X, Y) - 1)}$, em que $\alpha > 0$, podem ser interpretadas como medidas de similaridade fortes, no sentido amplo de De Baets e De Meyer, se o universo de discurso for uma malha $M \times N$, porque κ e σ são simétricas e satisfazem $\kappa(A, B) = 1 \iff A = B$ e $\sigma(A, B) = 1 \iff A = B$.

No próximo capítulo, apresentaremos uma descrição dos principais modelos da literatura que motivaram o desenvolvimento desta tese.

3 Memórias Associativas

Memórias associativas são modelos matemáticos cujo objetivo principal é armazenar um conjunto de pares $\{(\mathbf{a}^\xi, \mathbf{b}^\xi), \xi = 1, \dots, p\}$ chamado de conjunto das memórias fundamentais e, como já mencionamos anteriormente, podem ser classificadas em memórias autoassociativas ou heteroassociativas. Lembremos que uma memória associativa é chamada autoassociativa se $\mathbf{a}^\xi = \mathbf{b}^\xi$ para todo $\xi \in \{1, \dots, p\}$ e é chamada heteroassociativa se existe $\xi \in \{1, \dots, p\}$ para o qual $\mathbf{a}^\xi \neq \mathbf{b}^\xi$ [26, 27].

Lembremos também que utilizaremos o termo *memórias associativas bipolares* para nos referir às memórias associativas projetadas para armazenamento e recordação de padrões bipolares (vetores com coordenadas -1 ou 1). Nos referiremos às memórias associativas projetadas para armazenamento e recordação de padrões reais (vetores com coordenadas reais) como memórias associativas reais. Por fim, recordemos que memórias associativas *fuzzy* são memórias associativas projetadas para armazenamento e recordação uma família finita de conjuntos *fuzzy* [26, 27, 74].

Neste capítulo, descreveremos alguns modelos da literatura que motivaram o desenvolvimento desta tese. Iniciaremos com modelos projetados para o armazenamento e recordação de vetores bipolares (ou binários), a saber: rede de Hopfield [34], as memórias associativas recorrentes por correlação (RCAMs) [9], as memórias associativas exponenciais por correlação (ECAM) [9] e as memórias associativas bidirecionais exponenciais (EBAMs) [37]. Em seguida, recordaremos as definições das memórias associativas recorrentes exponenciais multivaloradas (MERAM) [10] e as memórias associativas com núcleo (KAM) [84], modelos que podem ser utilizados para o armazenamento e recordação de vetores multivalorados ou reais. Por fim, apresentaremos as memórias associativas recorrentes exponenciais *fuzzy* (RE-FAMs) [75], as memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs) [64, 70] e as Θ -FAMs [17], que são modelos de memórias associativas *fuzzy*.

Ao longo dos próximos capítulos, quando for necessário, escreveremos $H = (h_{ij})_{m \times n}$ para especificar que o elemento da i -ésima linha e j -ésima coluna de uma determinada matriz H será denotado por h_{ij} e enfatizar que matriz M tem dimensão $m \times n$.

3.1 Memórias Associativas Bipolares

A rede (discreta) de Hopfield é uma rede neural artificial recorrente que ganhou destaque após a publicação do trabalho “*Neural Networks and Physical Systems with*

Emergent Collective Computational Abilities” em 1982 por John Hopfield [34]. Um dos modelos mais conhecidos utilizados para implementar uma memória autoassociativa para armazenamento e recordação de padrões bipolares, a rede de Hopfield é composta por uma única camada com n neurônios de McCulloch e Pitts [29, 50] totalmente conectados. Cada neurônio possui dois estados possíveis, a saber -1 ou 1 (ou 0 e 1).

Sejam $\mathcal{A} = \{\mathbf{a}^1, \dots, \mathbf{a}^p\} \subset \{-1, 1\}^n$ o conjunto das memórias fundamentais e $\mathbf{x}_t = (x_{t1}, \dots, x_{tn}) \in \{-1, 1\}^n$ o vetor cuja ξ -ésima coordenada, $x_{t\xi}$, representa o estado do ξ -ésimo neurônio no instante t . Dado um vetor de entrada $\mathbf{x}_0 \in \{-1, 1\}^n$, a evolução da rede de Hopfield é descrita pela equação:

$$x_{t+1i} = \begin{cases} 1, & \text{se } \sum_{j=1}^n w_{ij}x_{tj} > T_i, \\ x_{ti}, & \text{se } \sum_{j=1}^n w_{ij}x_{tj} = T_i, \\ -1, & \text{se } \sum_{j=1}^n w_{ij}x_{tj} < T_i, \end{cases} \quad (3.1)$$

para $i = 1, \dots, n$, onde a matriz dos pesos sinápticos $W = (w_{ij})_{n \times n}$ é definida por

$$w_{ij} = \begin{cases} \sum_{k=1}^p a_i^k a_j^k, & \text{se } i \neq j, \\ 0, & \text{se } i = j \end{cases} \quad (3.2)$$

e T_i , $i = 1, \dots, n$ são valores limiares (threshold) definidos para cada neurônio. Consideraremos $T_i = 0$, para todo $i = 1, \dots, n$, conforme sugerido por Hopfield [34].

O elemento w_{ij} da matriz W indica a “força” da conexão entre os neurônios j e i . A definição da matriz W é baseada na aprendizagem de Hebb [29, 30]. Assim, para cada memória fundamental \mathbf{a}^k , a igualdade das componentes i e j de \mathbf{a}^k contribui para o aumento do peso sináptico w_{ij} , enquanto que, componentes i e j de \mathbf{a}^k diferentes contribuem para um peso sináptico w_{ij} menor. Além disso, a matriz W é simétrica, isto é, a força da conexão do neurônio i com o neurônio j é igual a força da conexão do neurônio j com o neurônio i . Ainda, temos que a matriz W possui diagonal principal nula, o que significa que os neurônios não estão conectados a si mesmos. A Figura 2 mostra um diagrama de blocos da rede de Hopfield. As cores no diagrama enfatizam que os neurônios não estão conectados a si mesmos e z^{-1} denota um atraso unitário no tempo.

A cada instante, apenas um neurônio atualiza seu estado conforme a equação de evolução (3.1). Em outras palavras, a atualização da rede deve ser feita de modo assíncrono. A atualização assíncrona dos neurônios garante a convergência da sequência produzida pelo modelo para um estado que minimiza uma função energia, conforme descrito em [34].

Apesar de suas diversas aplicações, a rede de Hopfield apresenta baixa capaci-

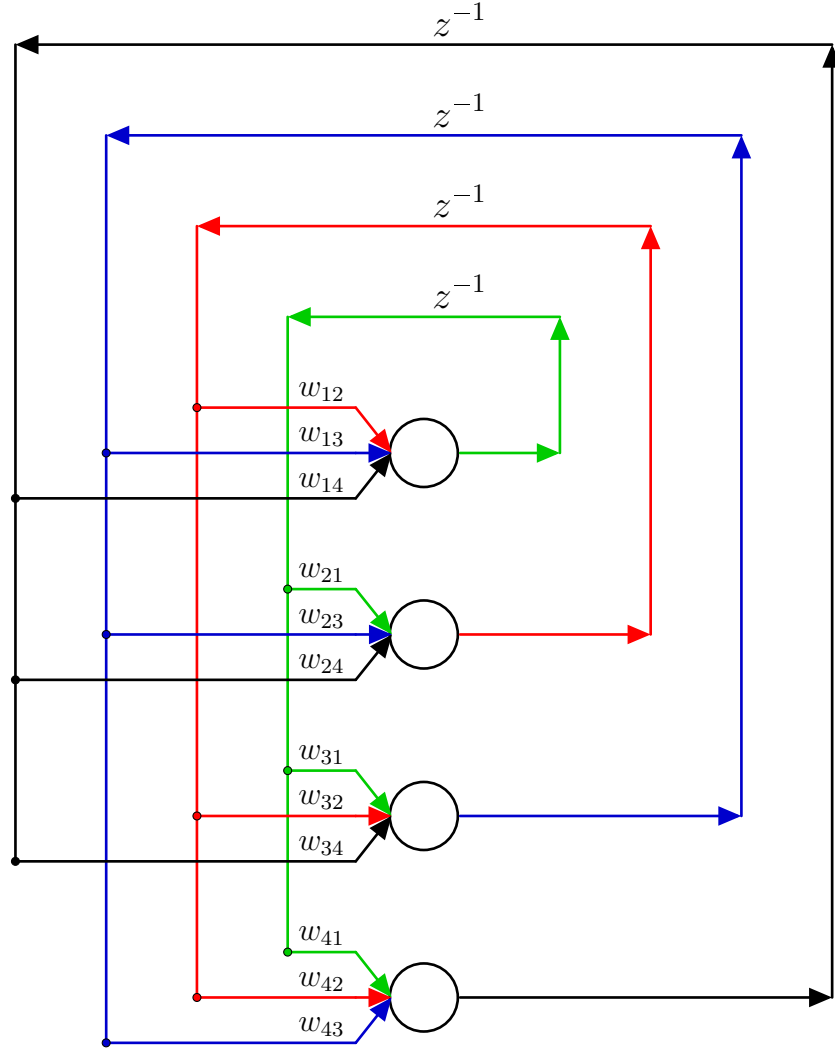


Figura 2 – Diagrama de blocos de uma rede de Hopfield com quatro neurônios.

dade de armazenamento [51]. Este fato motivou pesquisadores a buscarem alternativas para superar essa limitação. Por exemplo, Chiueh e Goodman propuseram uma generalização da rede de Hopfield por meio do acréscimo de uma camada intermediária de neurônios, a qual calcula uma medida de correlação entre o padrão de estado atual da rede e os itens armazenados. O modelo obtido, denominado memórias associativas recorrentes por correlação (RCAM, do inglês *Recurrent Correlation Associative Memories*) [9], apresenta uma melhora significativa na capacidade de armazenamento se comparado a rede de Hopfield. As RCAMs são definidas da seguinte forma.

Seja $\mathcal{A} = \{\mathbf{a}^1, \dots, \mathbf{a}^p\} \subset \{-1, 1\}^n$ o conjunto das memórias fundamentais. Dado um vetor de entrada $\mathbf{x}_0 \in \{-1, 1\}^n$, a dinâmica de uma memória associativa recorrente por correlação é descrita pela equação:

$$\mathbf{x}_{t+1} = \text{sgn} \left(\sum_{j=1}^p f_j(\langle \mathbf{a}^j, \mathbf{x}_t \rangle) \mathbf{a}^j \right), \quad (3.3)$$

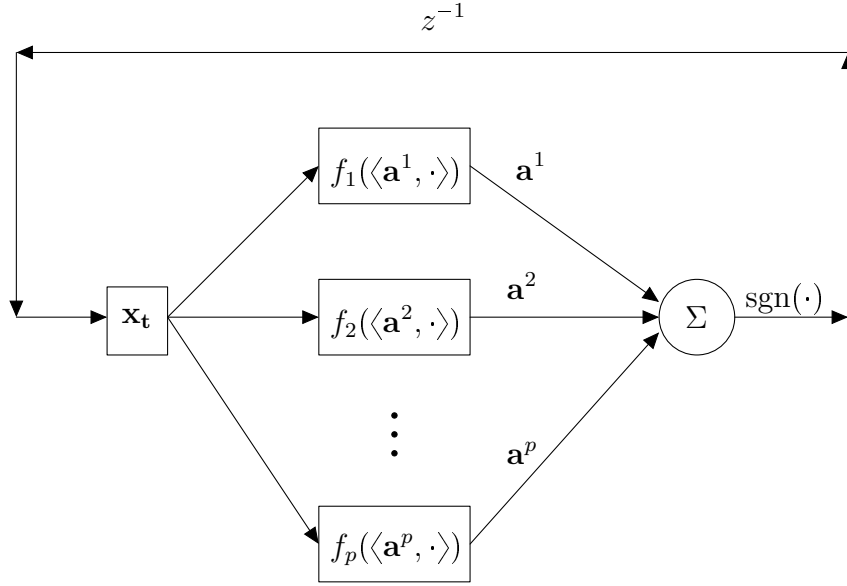


Figura 3 – Diagrama de blocos da RCAM

em que $f_j : [-n, n] \rightarrow \mathbb{R}$, $j = 1, \dots, p$ são chamadas funções peso, $\langle \cdot, \cdot \rangle$ denota o produto interno usual e a função sinal $\text{sgn} : \mathbb{R}^n \rightarrow \{-1, 1\}^n$ é calculada componente-a-componente da seguinte forma:

$$x_{t+1i} = \text{sgn} \left(\sum_{j=1}^p f_j(\langle \mathbf{a}^j, \mathbf{x}_t \rangle) a_i^j \right) = \begin{cases} 1, & \text{se } \sum_{j=1}^p f_j(\langle \mathbf{a}^j, \mathbf{x}_t \rangle) a_i^j > 0 \\ x_{ti}, & \text{se } \sum_{j=1}^p f_j(\langle \mathbf{a}^j, \mathbf{x}_t \rangle) a_i^j = 0 \\ -1, & \text{se } \sum_{j=1}^p f_j(\langle \mathbf{a}^j, \mathbf{x}_t \rangle) a_i^j < 0. \end{cases}$$

Especificamente, uma RCAM é uma rede neural recorrente com duas camadas intermediárias. A primeira camada, composta por p neurônios, calcula uma medida de correlação entre o estado atual da rede \mathbf{x}_t e as memórias fundamentais. A segunda camada, por sua vez, utiliza os pesos fornecidos pela camada anterior para produzir uma soma ponderada das memórias fundamentais. Por fim, aplica-se a função sinal para garantir que a condição $\mathbf{x}_t \in \{-1, 1\}^n$ seja válida para $t > 0$. A Figura 3 mostra um diagrama de blocos da RCAM. Na Figura 3, z^{-1} denota um atraso unitário no tempo.

Conforme demonstrado em [9], a RCAM definida pela equação (3.3) converge em ambos os modos síncrono e assíncrono de atualização se considerarmos funções peso $f_j = f$, para todo $j = 1 : \dots, p$, em que $f : [-n, n] \rightarrow \mathbb{R}$ é contínua e monótona não decrescente. Por serem não decrescentes, as funções peso enfatizam a contribuição das memórias fundamentais no processo de recordação, de modo que o peso da memória fundamental que possui maior correlação com \mathbf{x}_t seja muito maior que os demais no cálculo de \mathbf{x}_{t+1} por meio a equação (3.3).

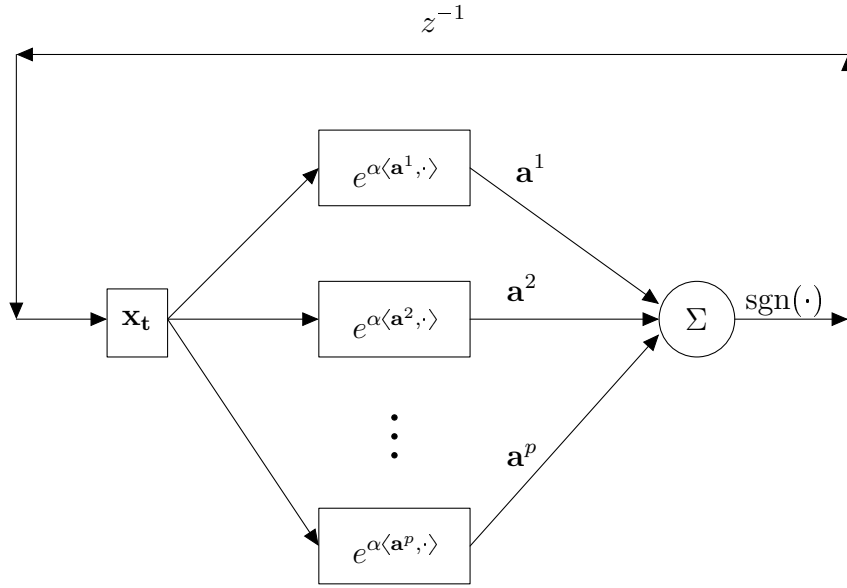


Figura 4 – Diagrama de blocos da ECAM

As memórias associativas exponenciais por correlação [9] (ECAMs, do inglês *Exponential Correlation Associative Memories*) são obtidas considerando em (3.3) funções peso $f_j : [-n, n] \rightarrow \mathbb{R}$ definidas, para todo $j = 1, \dots, p$, por:

$$f_j(t) = e^{\alpha t}, \quad \alpha > 0. \quad (3.4)$$

Especificamente, as ECAMs são definidas da seguinte forma:

Definição 13 (ECAM). *Seja $\{\mathbf{a}^\xi, \xi = 1, \dots, p\} \subset \{-1, 1\}^n$ um conjunto de memórias fundamentais e $\alpha > 0$ um número real. Dado um padrão de entrada $\mathbf{x}_0 \in \{-1, 1\}^n$, a ECAM produz uma sequência de vetores $\mathbf{x}_t \in \{-1, 1\}^n$ definidos por:*

$$\mathbf{x}_{t+1} = \text{sgn} \left\{ \sum_{\xi=1}^p e^{\alpha \langle \mathbf{a}^\xi, \mathbf{x}_t \rangle} \mathbf{a}^\xi \right\}, \quad \forall t > 0, \quad (3.5)$$

em que $\langle \mathbf{a}^\xi, \mathbf{x}_t \rangle$ denota o produto interno dos vetores \mathbf{a}^ξ e \mathbf{x}_t e sgn é a função sinal.

A Figura 4 apresenta um diagrama de blocos da ECAM. Na Figura 4, z^{-1} denota um atraso unitário no tempo.

Observemos que, como o produto interno entre vetores bipolares $\mathbf{x}, \mathbf{y} \in \{-1, 1\}^n$ satisfaz a identidade:

$$\langle \mathbf{x}, \mathbf{y} \rangle = n - \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2, \quad (3.6)$$

podemos interpretá-lo, de forma intuitiva, como uma medida da similaridade entre os vetores \mathbf{x} e \mathbf{y} .

Além da excelente tolerância a ruídos, as ECAMs apresentam grande capacidade de armazenamento. De fato, a capacidade de armazenamento das ECAMs cresce exponencialmente com a dimensão dos padrões \mathbf{a}^ξ , $\xi = 1, \dots, p$. Especificamente, c^n padrões de dimensão n podem ser armazenados na ECAM, em que $c \geq 1$ [9].

Para o caso heteroassociativo, Jeng et al. introduziram as memórias associativas bidirecionais exponenciais (EBAMs, do inglês *Exponential Bidirectional Associative Memories*) [37], também projetada para o armazenamento de pares de vetores bipolares [37]. A EBAM pode ser vista como uma versão heteroassociativa da ECAM. Formalmente, as EBAMs são definidas da seguinte maneira:

Definição 14 (EBAM). *Sejam $\{(\mathbf{a}^\xi, \mathbf{b}^\xi) \mid \xi = 1, \dots, p\} \subset \{-1, 1\}^n \times \{-1, 1\}^m$ um conjunto de memórias fundamentais e $\alpha > 0$ um número real. Dado um padrão de entrada $\mathbf{x}_0 \in \{-1, 1\}^n$, a EBAM produz recursivamente sequências de vetores $\mathbf{x}_t \in \{-1, 1\}^n$, $\mathbf{y}_t \in \{-1, 1\}^m$, definidos por:*

$$\mathbf{y}_t = \text{sgn} \left\{ \sum_{\xi=1}^p e^{\alpha \langle \mathbf{a}^\xi, \mathbf{x}_t \rangle} \mathbf{b}^\xi \right\} \quad e \quad \mathbf{x}_{t+1} = \text{sgn} \left\{ \sum_{\xi=1}^p e^{\alpha \langle \mathbf{b}^\xi, \mathbf{y}_t \rangle} \mathbf{a}^\xi \right\}, \forall t > 0. \quad (3.7)$$

A Figura 5 mostra um diagrama de blocos da EBAM. Na Figura 5, z^{-1} denota um atraso unitário no tempo.

A ECAM e a EBAM são memórias associativas projetadas para o armazenamento e recordação de vetores bipolares. No entanto, em algumas aplicações as memórias fundamentais são vetores com coordenadas reais ou conjuntos *fuzzy*. Na próxima seção, recordaremos a definição de memórias associativas propostas para o armazenamento de vetores com componentes reais.

3.2 Memórias Associativas Reais

As memórias associativas recorrentes multivaloradas (MERAM, do inglês *Multivalued Exponential Recurrent Associative Memory*) [10], introduzidas por Chiueh e Tsai, estendem as memórias associativas recorrentes por correlação para o armazenamento e recordação de vetores multivalorados. Para simplificar a exposição, apresentaremos as MERAMs para o armazenamento e recordação de vetores com valores reais.

Consideremos $\{\mathbf{a}^1, \dots, \mathbf{a}^p\} \subset \mathbb{R}^n$ o conjunto das memórias fundamentais. Dado um padrão de entrada \mathbf{x}_0 , as MERAMs são definidas por meio da seguinte equação de evolução:

$$\mathbf{x}_{t+1} = \frac{\sum_{\xi=1}^p e^{\alpha \Psi(\mathbf{a}^\xi, \mathbf{x}_t)} \mathbf{a}^\xi}{\sum_{\eta=1}^p e^{\alpha \Psi(\mathbf{a}^\eta, \mathbf{x}_t)}}, \quad (3.8)$$

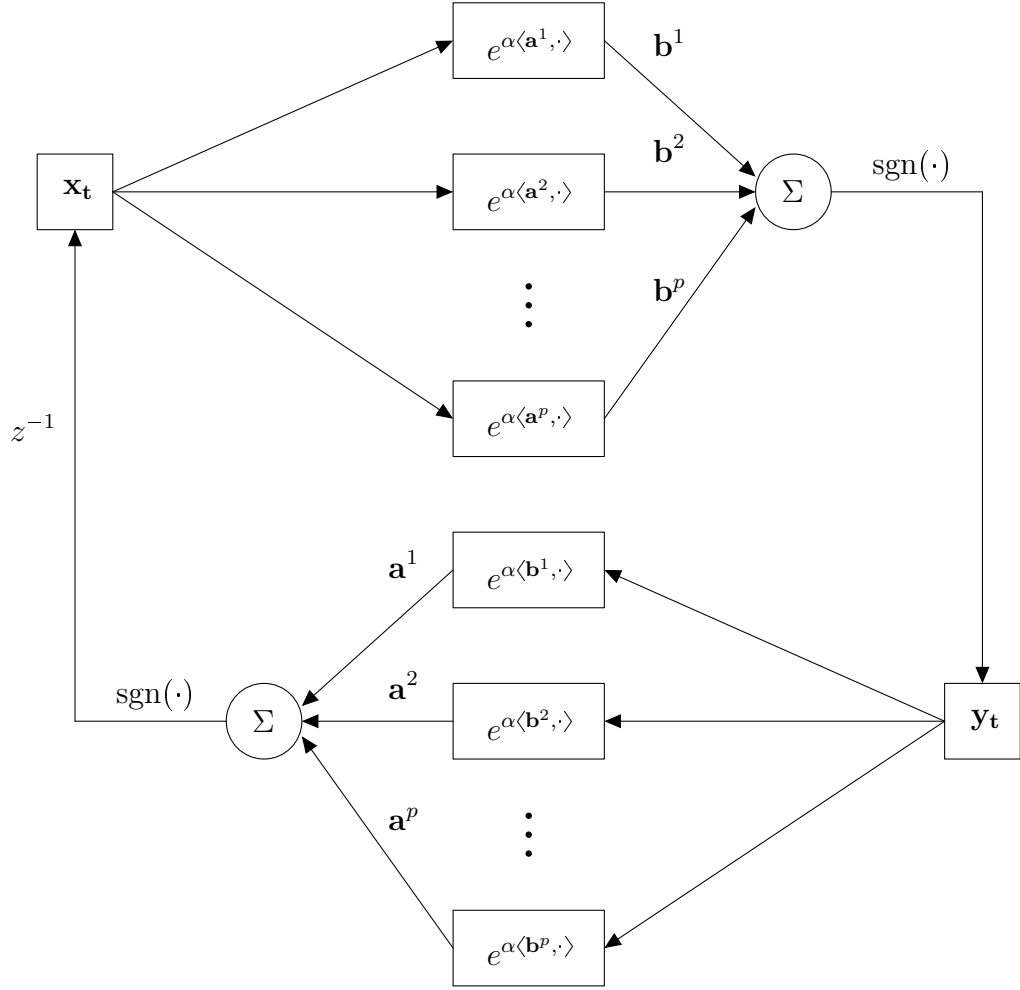


Figura 5 – Diagrama de blocos da EBAM.

onde $\alpha > 0$ é uma constante e Ψ denota uma medida de similaridade em um sentido amplo.

A Figura 6 apresenta um diagrama de blocos da MERAM. Na Figura 6, z^{-1} denota um atraso unitário no tempo. Além disso, lembremos que uma combinação linear de vetores $\sum_{i=1}^p w_i \mathbf{x}^i$ é chamada combinação afim se $\sum_{i=1}^p w_i = 1$.

Como no caso da ECAM, a função exponencial contribui para que as memórias fundamentais mais similares ao padrão \mathbf{x}_t , segundo a medida Ψ , tenham um peso maior no cálculo de \mathbf{x}_{t+1} , $t > 0$. Isto pode ajudar a melhorar a capacidade de armazenamento e recordação da rede [10].

Outros modelos de memória associativa para armazenamento de vetores reais são as memórias associativas com núcleo (KAMs, do inglês *Kernel Associative Memories*). As KAMs foram propostas por Zhang et al [84] como uma extensão das memórias associativas lineares por correlação, conforme descreveremos a seguir.

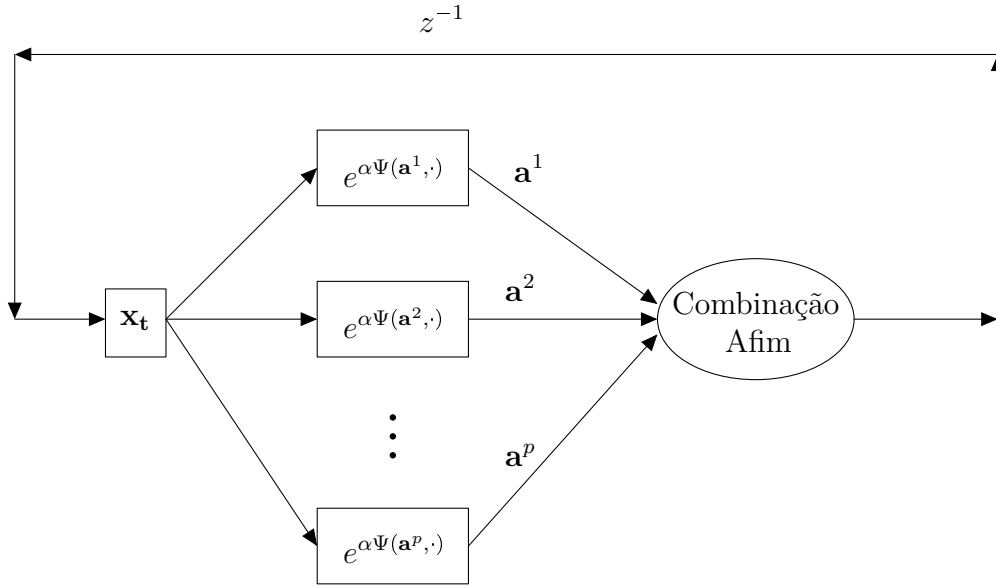


Figura 6 – Diagrama de blocos da MERAM.

As memórias associativas lineares [5, 40, 53] são caracterizadas pelo fato de que a aplicação que representa a memória é uma função linear. Dessa forma, considerando $\{(\mathbf{a}^1, \mathbf{b}^1), \dots, (\mathbf{a}^p, \mathbf{b}^p)\} \subset \mathbb{R}^n \times \mathbb{R}^m$ o conjunto das memórias fundamentais, o objetivo é determinar uma matriz M tal que $M\mathbf{a}^\xi = \mathbf{b}^\xi$, para todo $\xi = 1, \dots, p$. Tal matriz M pode ser construída por meio de uma técnica chamada de armazenamento por correlação, conforme a seguinte expressão:

$$M = \sum_{\xi=1}^p \mathbf{b}^\xi (\mathbf{a}^\xi)^T = B A^T, \quad (3.9)$$

onde A é a matriz cuja ξ -ésima coluna é o vetor \mathbf{a}^ξ , B é a matriz cuja ξ -ésima coluna é \mathbf{b}^ξ e $(\mathbf{a}^\xi)^T$, A^T denotam, respectivamente, o vetor \mathbf{a}^ξ e a matriz A transpostos.

As memórias associativas lineares definidas utilizando a matriz M obtida por meio do armazenamento por correlação são chamadas de memórias associativas lineares por correlação. Em termos matemáticos, dado um padrão de entrada \mathbf{x} , as memórias associativas lineares por correlação produzem um vetor \mathbf{y} dado por:

$$\mathbf{y} = M\mathbf{x} = \sum_{\xi=1}^p \mathbf{b}^\xi (\mathbf{a}^\xi)^T \mathbf{x} = \sum_{\xi=1}^p \langle \mathbf{a}^\xi, \mathbf{x} \rangle \mathbf{b}^\xi, \quad (3.10)$$

onde $\langle \mathbf{a}^\xi, \mathbf{x} \rangle$ denota o produto interno usual entre os vetores \mathbf{a}^ξ e \mathbf{x} .

No entanto, as memórias associativas lineares por correlação apresentam baixa capacidade absoluta de armazenamento. Visando superar essa restrição, Zhang et al. propuseram uma modificação nos coeficientes da combinação linear (3.10), substituindo-os

por uma função núcleo [84]. Por meio dessa substituição, obtemos:

$$\mathbf{y} = \sum_{\xi=1}^p k(\mathbf{a}^\xi, \mathbf{x}) \mathbf{b}^\xi, \quad (3.11)$$

onde $k(\cdot, \cdot)$ é uma função núcleo [84].

Um exemplo muito conhecido de função núcleo é a função de base radial Gaussiana [31] definida por:

$$\mathbf{k}(\mathbf{x}, \mathbf{y}) = \exp \left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2} \right). \quad (3.12)$$

O modelo dado por (3.11) ainda pode sofrer com uma baixa capacidade de armazenamento devido à interferência cruzada dos itens memorizados. Com o objetivo de evitar esse problema, as memórias associativas com núcleo (KAM, *kernel associative memory*) são definidas por meio da generalização de (3.11) para a forma paramétrica:

$$\mathbf{y} = \sum_{\xi=1}^p \mathbf{w}^\xi k(\mathbf{a}^\xi, \mathbf{x}). \quad (3.13)$$

Os vetores \mathbf{w}^i , $i = 1, \dots, p$ minimizam o erro

$$\sum_{\xi=1}^p \|\mathbf{b}^\xi - W\mathbf{k}^\xi\|^2, \quad (3.14)$$

em que W denota a matriz cuja i -ésima coluna é o vetor \mathbf{w}^i e \mathbf{k}^ξ denota o vetor p -dimensional cuja j -ésima coordenada é dada por $k_j^\xi = k(\mathbf{a}^\xi, \mathbf{a}^j)$ [84, 86]. Assim, a matriz W é definida por $W = BK^\dagger$, onde B é a matriz cujas colunas são os vetores \mathbf{b}^i , K é a matriz cujas colunas são os vetores \mathbf{k}^i , e K^\dagger denota a pseudo-inversa da matriz K .

Em [84], Zhang et al propuseram a seguinte versão normalizada da KAM:

$$\mathbf{y} = \frac{\sum_{\xi=1}^p \hat{\mathbf{w}}^\xi k(\mathbf{a}^\xi, \mathbf{x})}{\sum_{\eta=1}^p k(\mathbf{a}^\eta, \mathbf{x})}, \quad (3.15)$$

onde $\hat{\mathbf{w}}^\xi$ é a ξ -ésima coluna da matriz $\widehat{W} = B\widehat{K}^\dagger$, em que B é a matriz cujas colunas são $B(:, \xi) = \mathbf{b}^\xi$, \widehat{K} é a matriz cujas colunas são os vetores \mathbf{k}^ξ normalizados, isto é, $\widehat{K}(:, \xi) = \frac{\mathbf{k}^\xi}{\|\mathbf{k}^\xi\|_1}$, $\xi = 1, \dots, p$ e \widehat{K}^\dagger denota a pseudo-inversa da matriz \widehat{K} . Observemos que $\mathbf{k}(\mathbf{x}, \mathbf{y})$ pode ser muito próximo de 0 no caso em que \mathbf{x} e \mathbf{y} são muito diferentes. Dessa forma, a normalização proposta contribui para que as componentes dos vetores \mathbf{k}^ξ , $\xi = 1, \dots, p$ não sejam todas muito próximas de 0.

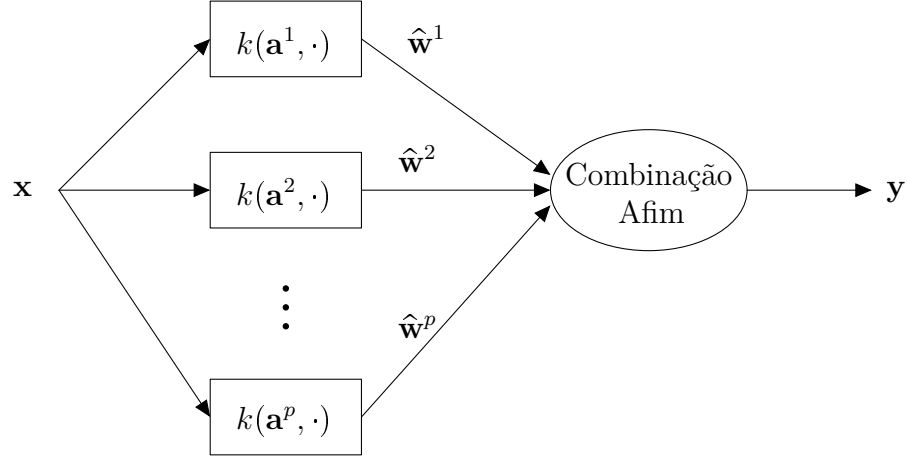


Figura 7 – Diagrama de blocos da KAM.

A Figura 7 mostra um diagrama de blocos da KAM normalizada. Na Figura 7, $\hat{\mathbf{w}}^\xi$ denota a ξ -ésima coluna da matriz $\widehat{W} = B\widehat{K}^\dagger$, definida após a equação (3.15).

Conforme descrito em [86], apesar de ter sido proposta como uma extensão das memórias associativas lineares por correlação, de certa forma, a KAM pode ser considerada um autoassociador com núcleo, e portanto, possui aplicações em classificação de padrões. Além disso, conforme [84], a KAM pode ser aplicada também a problemas de reconhecimento de faces.

3.3 Memórias Associativas Fuzzy

As memórias associativas recorrentes exponenciais *fuzzy* (RE-FAMs, do inglês *Recurrent Exponential Fuzzy Associative Memories*), introduzidas por Valle [75], são modelos autoassociativos projetados para armazenamento e recordação de uma família finita de conjuntos *fuzzy* $\mathcal{A} = \{A^1, \dots, A^p\}$ e podem ser vistas como uma versão *fuzzy* da ECAM.

Precisamente, uma RE-FAM é definida da seguinte forma [75]:

Definição 15 (RE-FAM). *Sejam $\{A^1, \dots, A^p\} \subset \mathcal{F}(U)$ o conjunto das memórias fundamentais, $\alpha > 0$ um número real e \mathcal{S} uma medida de similaridade. Dado um conjunto fuzzy $X_0 \in \mathcal{F}(U)$, as RE-FAMs produzem recursivamente uma sequência de conjuntos fuzzy $X_t \in \mathcal{F}(U)$ definidos, para todo $u \in U$ e $t = 0, 1, \dots$, por:*

$$X_{t+1}(u) = \sum_{i=1}^p w_{ti} A^i(u), \quad (3.16)$$

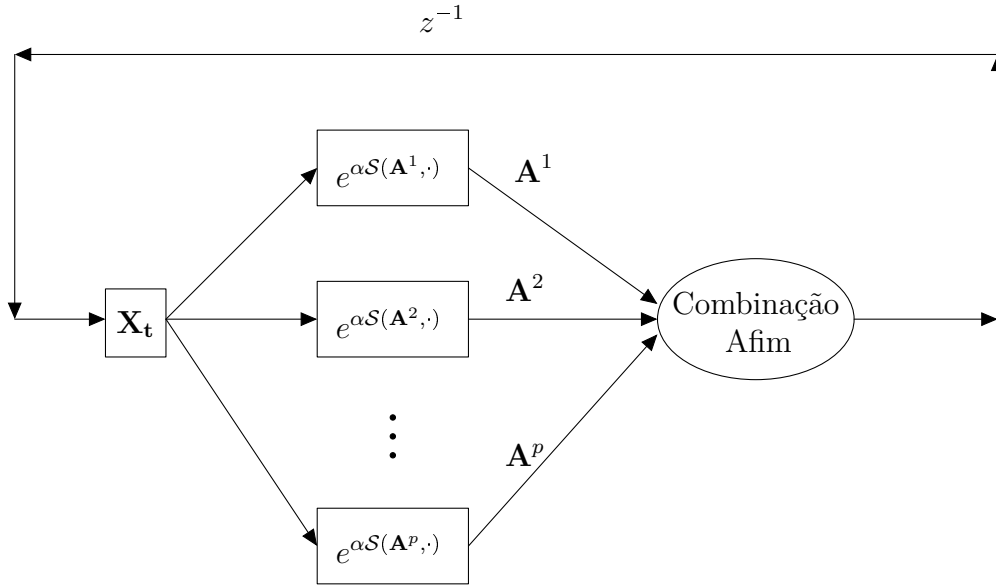


Figura 8 – Diagrama de blocos da REFAM.

em que os pesos w_{ti} são calculados, para todo $i = 1, \dots, p$ e $t \geq 0$ por meio da equação:

$$w_{ti} = \frac{e^{\alpha \mathcal{S}(A^i, X_t)}}{\sum_{j=1}^p e^{\alpha \mathcal{S}(A^j, X_t)}}.$$

Observemos que X_{t+1} é obtido por meio de uma combinação linear das memórias fundamentais cujos coeficientes são calculados utilizando uma exponencial de uma medida de similaridade.

Assim como as ECAMs, as RE-FAMs são redes neurais recorrentes com duas camadas. A primeira camada calcula uma exponencial da medida de similaridade entre as memórias fundamentais A^ξ , $\xi = 1, \dots, p$ e o vetor X_t . A segunda camada, utilizando as saídas da camada anterior, produz uma combinação afim dos conjuntos *fuzzy* A^1, \dots, A^p . Lembremos que X_{t+1} é uma combinação afim de A^ξ , $\xi = 1, \dots, p$ se

$$X_{t+1} = \sum_{i=1}^p w_{ti} A^i \text{ com } \sum_{i=1}^p w_{ti} = 1.$$

A Figura 8 apresenta um diagrama de blocos da RE-FAM.

A saída de um único passo da RE-FAM converge para a média aritmética das memórias fundamentais mais similares ao padrão de entrada quando o parâmetro α tende a infinito [64]. No entanto, se o parâmetro α não for suficientemente grande, o conjunto das memórias fundamentais pode não ser corretamente armazenado na RE-FAM, devido à interferência cruzada das memórias fundamentais [64]. Este fato motivou o surgimento

das memórias associativas recorrentes exponenciais *fuzzy* generalizadas [64, 70], modelo que, sob certas condições, é capaz de armazenar corretamente o conjunto das memórias fundamentais para qualquer valor do parâmetro α .

As memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs, do inglês *Generalized Recurrent Exponential Fuzzy Associative Memories*) generalizam as RE-FAMs por meio do acréscimo de uma camada oculta de neurônios lineares com o objetivo de reduzir a interferência cruzada entre os itens memorizados [64, 70]. Especificamente, dado um conjunto *fuzzy* inicial $X_0 \in \mathcal{F}(U)$, as GRE-FAMs produzem uma sequência de conjuntos *fuzzy* $X_t \in \mathcal{F}(U)$ a partir de uma combinação afim das memórias fundamentais conforme a definição a seguir [64, 70]:

Definição 16 (GRE-FAM). *Sejam $\mathcal{A} = \{A^1, A^2, \dots, A^p\} \subseteq \mathcal{F}(U)$ uma família finita de memórias fundamentais, $\alpha > 0$ um número real e $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ uma medida de similaridade. Consideremos $C = (c_{ij})$ a matriz definida por $c_{ij} = e^{\alpha \mathcal{S}(A^i, A^j)}$, para todo $i, j = 1, \dots, p$ e defina $G = C^{-1}$. Dado um conjunto *fuzzy* de entrada, $\mathbf{x}_0 \in \mathcal{F}(U)$, a GRE-FAM produz uma sequência de conjuntos *fuzzy* $\{X_t\}_{t \geq 0}$ definidos, para quaisquer $u \in U$ e $t = 0, 1, \dots$, por:*

$$X_{t+1}(u) = \varphi \left(\sum_{\xi=1}^p w_{\xi t} A^\xi(u) \right), \quad (3.17)$$

em que os pesos $w_{\xi t}$ são calculados por meio da expressão

$$w_{\xi t} = \frac{\sum_{\mu=1}^p g_{\xi \mu} e^{\alpha \mathcal{S}(A^\mu, X_t)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta \mu} e^{\alpha \mathcal{S}(A^\mu, X_t)}} \quad (3.18)$$

e a função $\varphi : \mathbb{R} \rightarrow [0, 1]$, definida por

$$\varphi(x) = \max\{0, \min\{1, x\}\}, \quad (3.19)$$

garante que $X_{t+1}(u) \in [0, 1]$, para quaisquer $t \geq 0$ e $u \in U$.

A Figura 9 apresenta um diagrama de blocos da GRE-FAM.

Observação 4. Na Definição 16, consideramos a matriz $G = C^{-1}$ pois, desta forma, todas as memórias fundamentais são corretamente armazenadas na GRE-FAM. Em outras palavras, as memórias fundamentais são pontos fixos da GRE-FAM quando $G = C^{-1}$ [64, 76]. Caso a matriz C não seja inversível, consideramos $G = C^\dagger$, em que C^\dagger denota a pseudo-inversa de C [23].

As GRE-FAMs são modelos de memórias autoassociativas, isto é, são projetadas para armazenar memórias fundamentais da forma (A^ξ, B^ξ) com $A^\xi = B^\xi$. No entanto,

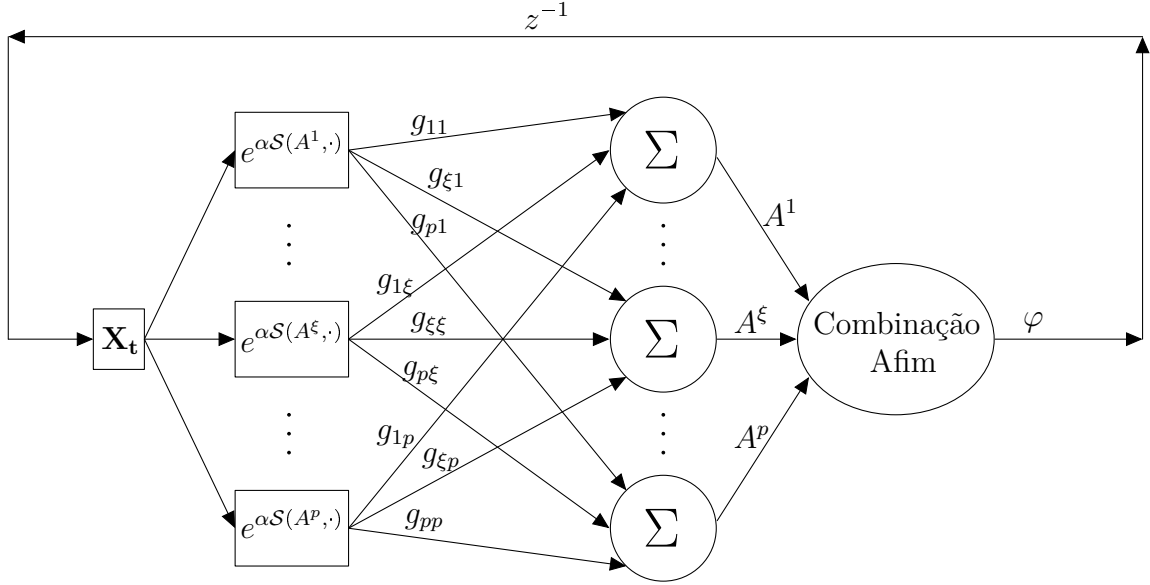


Figura 9 – Diagrama de blocos da GREFAM.

algumas aplicações requerem o armazenamento de pares (A^{ξ}, B^{ξ}) em que $A^{\xi} \neq B^{\xi}$. No próximo capítulo, apresentaremos a definição da memória associativa bidirecional exponencial *fuzzy* generalizada, modelo inspirado pela EBAM [37] que generaliza as GRE-FAMs para o caso heteroassociativo.

As memórias associativas Θ -Fuzzy (Θ -FAMs, do inglês *Θ -Fuzzy Associative Memories*), introduzidas por Esmi et al [17], são uma classe de memórias associativas *fuzzy* parametrizadas. Em termos matemáticos, as Θ -FAMs são definidas conforme descrito a seguir.

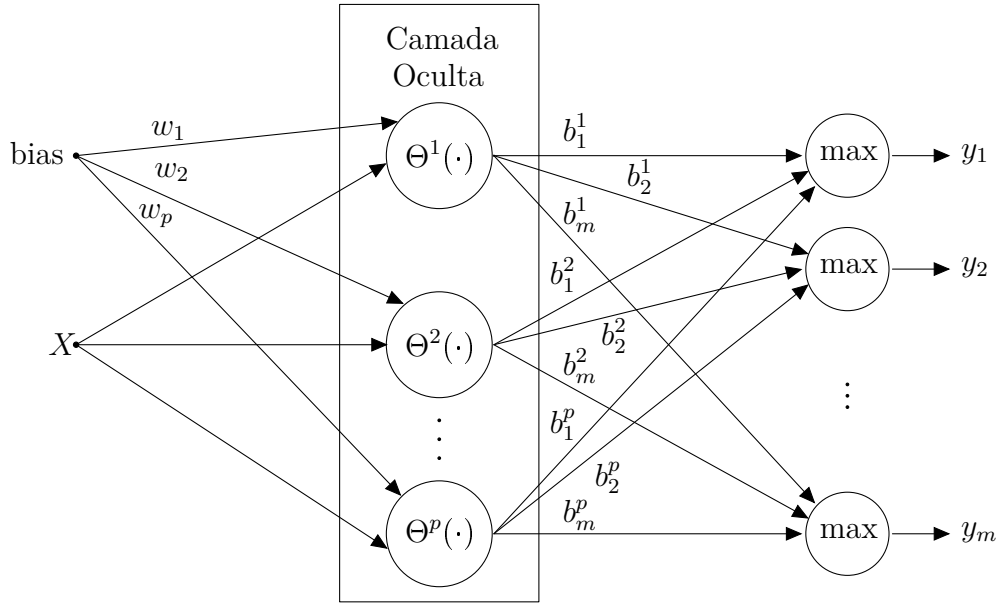
Sejam $\{(A^{\xi}, B^{\xi}), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais, $W \subset \mathbb{R}^p$ um conjunto fechado convexo e $\Theta^{\xi} : \mathcal{F}(U) \rightarrow [0, 1]$ operadores tais que $\Theta^{\xi}(A^{\xi}) = 1$, para $\xi = 1, \dots, p$. Dados um conjunto *fuzzy* $X \in \mathcal{F}(U)$ e um vetor $\mathbf{w} = (w_1, w_2, \dots, w_p) \in W$, a Θ -FAM [17] baseada em Θ^{ξ} e \mathbf{w} produz o conjunto *fuzzy* $Y = \mathcal{T}(X) \in \mathcal{F}(V)$ definido por:

$$Y = \mathcal{T}(X) = \bigcup_{j \in I_{\mathbf{w}}(X)} B^j,$$

em que $I_{\mathbf{w}}(X)$ é o conjunto de índices definido por:

$$I_{\mathbf{w}}(X) = \left\{ j \in \{1, \dots, p\} : w_j \Theta^j(X) = \max_{\xi=1, \dots, p} w_{\xi} \Theta^{\xi}(X) \right\}.$$

Observemos que as Θ -FAMs possuem uma camada oculta baseada em funções Θ^{ξ} formada por p neurônios. O ξ -ésimo neurônio da camada oculta calcula $w_{\xi} \Theta^{\xi}(X)$ e produz 1 como saída se, e somente se, $w_{\xi} \Theta^{\xi}(X) \geq w_{\eta} \Theta^{\eta}(X)$, para todo $\xi \neq \eta$. A Figura 10 mostra um diagrama de blocos da Θ -FAM.

Figura 10 – Diagrama de blocos da Θ -FAM.

Dado um conjunto de entrada $X \in \mathcal{F}(U)$, se existe um único índice ξ tal que $w_\xi \Theta^\xi(X) > w_\eta \Theta^\eta(X)$, para todo $\eta = 1, \dots, p$, $\eta \neq \xi$, então $I_{\mathbf{w}}(X) = \{\xi\}$ e portanto, a Θ -FAM produz $\mathcal{T}(X) = \bigcup_{j \in I_{\mathbf{w}}(X)} B^j = B^\xi$. Além disso, no caso autoassociativo, isto é, quando $B^\xi = A^\xi$, para todo $\xi = 1, \dots, p$, dado um conjunto *fuzzy* de entrada X , a Θ -FAM produz como saída o seguinte conjunto *fuzzy*:

$$\mathcal{T}(X) = \bigcup_{j \in I_{\mathbf{w}}(X)} A^j.$$

As Θ -FAMs são baseadas em funções $\Theta^\xi : \mathcal{F}(X) \rightarrow [0, 1]$. Tais funções, podem ser definidas a partir de medidas de similaridade. De fato, se $\mathcal{S} : \mathcal{F}(X) \times \mathcal{F}(X) \rightarrow [0, 1]$ é uma medida de similaridade que satisfaz a propriedade $\mathcal{S}(A, A) = 1$, para todo conjunto *fuzzy* $A \in \mathcal{F}(X)$, podemos definir funções Θ^ξ a partir de \mathcal{S} da seguinte forma: $\Theta^\xi(\cdot) = \mathcal{S}(A^\xi, \cdot)$, para todo $A \in \mathcal{F}(X)$. Uma Θ -FAM definida utilizando funções Θ^ξ obtidas a partir de medidas de similaridade são denominadas *SM-FAMs ponderadas* (do inglês *weighted similarity measure fuzzy associative memory*) [17, 19]. A Figura 10, extraída de [64], mostra um diagrama de blocos da Θ -FAM.

As TE-FAMs (*Tunable Equivalence Fuzzy Associative Memories*) [18], memórias associativas *fuzzy* pertencentes à classe das Θ -FAMs, são obtidas definindo $\Theta^\xi(\cdot) = E^\xi(\mathbf{x}^\xi, \cdot)$, em que E^ξ é uma medida de equivalência parametrizada, conforme descrito em [18].

4 Memórias Associativas Bidirecionais Exponenciais Fuzzy Generalizadas

No capítulo anterior, recordamos a definição das GRE-FAMs, memória autoassociativa projetada para armazenamento e recordação de conjuntos fuzzy [64, 70]. Neste capítulo motivados pelo fato de que algumas aplicações de memórias associativas requerem o armazenamento e recordação de um conjunto de pares (A^ξ, B^ξ) , com $A^\xi \neq B^\xi$, apresentaremos a definição das memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs, do inglês *generalized exponential bidirectional fuzzy associative memories*) bem como alguns resultados teóricos sobre este modelo. As GEB-FAMs constituem uma das principais contribuições desta tese e foram desenvolvidas com o objetivo de estender a definição das memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs) para o caso heteroassociativo.

Especificamente, as memórias associativas bidirecionais exponenciais *fuzzy* generalizadas (GEB-FAMs) são modelos matemáticos inspirados pela EBAM de Jeng et al. [37] e projetados para o armazenamento e recordação de uma família finita de pares de conjuntos *fuzzy*. Formalmente, a definição de uma GEB-FAM é dada por:

Definição 17 (GEB-FAM). *Seja $\{(A^\xi, B^\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais, $\alpha, \beta > 0$ números reais, $\mathcal{S}_U : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ e $\mathcal{S}_V : \mathcal{F}(V) \times \mathcal{F}(V) \rightarrow [0, 1]$ medidas de similaridade, G e H matrizes de dimensão $p \times p$. Dado um conjunto fuzzy de entrada $X_0 \in \mathcal{F}(U)$, uma GEB-FAM produz recursivamente sequências de conjuntos fuzzy $\{X_t\}_{t \geq 0}$ e $\{Y_t\}_{t \geq 0}$ definidos, para todo $t \geq 0$, $u \in U$ e $v \in V$, pelas seguintes equações:*

$$Y_t(v) = \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\mu=1}^p g_{\xi\mu} e^{\alpha \mathcal{S}_U(A^\mu, X_t)} B^\xi(v)}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} e^{\alpha \mathcal{S}_U(A^\mu, X_t)}} \right) \quad e \quad X_{t+1}(u) = \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\mu=1}^p h_{\xi\mu} e^{\beta \mathcal{S}_V(B^\mu, Y_t)} A^\xi(u)}{\sum_{\eta=1}^p \sum_{\mu=1}^p h_{\eta\mu} e^{\beta \mathcal{S}_V(B^\mu, Y_t)}} \right), \quad (4.1)$$

em que a função $\varphi : \mathbb{R} \rightarrow [0, 1]$, dada por $\varphi(x) = \max(0, \min(1, x))$, garante que $Y_t(v), X_{t+1}(u) \in [0, 1]$, para todo $v \in V$, $u \in U$ e para todo $t \geq 0$.

De maneira equivalente, podemos escrever as equações em (4.1), da seguinte

forma:

$$Y_t(v) = \varphi \left(\sum_{\xi=1}^p w_{\xi t}^A B^\xi(v) \right), \text{ em que } w_{\xi t}^A = \frac{\sum_{\mu=1}^p g_{\xi\mu} e^{\alpha S_U(A^\mu, X_t)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} e^{\alpha S_U(A^\mu, X_t)}} \quad (4.2)$$

e

$$X_{t+1}(u) = \varphi \left(\sum_{\xi=1}^p w_{\xi t}^B A^\xi(u) \right), \text{ em que } w_{\xi t}^B = \frac{\sum_{\mu=1}^p h_{\xi\mu} e^{\beta S_V(B^\mu, Y_t)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p h_{\eta\mu} e^{\beta S_V(B^\mu, Y_t)}}. \quad (4.3)$$

A equação (4.2) nos mostra que, a menos da função φ , Y_t é determinado, para todo $t \geq 0$, por meio de uma soma ponderada dos conjuntos fuzzy B^ξ em que os pesos $w_{\xi t}^A$ são calculados com base em uma exponencial das similaridades entre os conjuntos fuzzy A^ξ e X_t . De modo semelhante, conforme a (4.3), X_{t+1} é produzido, para todo $t \geq 0$, por meio de uma média ponderada dos A^ξ cujos pesos $w_{\xi t}^B$ são calculados a partir de uma exponencial das similaridades entre os conjuntos fuzzy B^ξ e Y_t .

A Figura 11 mostra um diagrama de blocos da GEB-FAM. No diagrama, Σ^{-1} simboliza a normalização realizada no cálculo dos pesos $w_{\xi t}^A$ e $w_{\xi t}^B$ e z^{-1} denota um atraso unitário no tempo.

Observemos que a GRE-FAM é um caso particular da GEB-FAM. De fato, no caso autoassociativo, isto é, quando o conjunto das memórias fundamentais $\{(A^\xi, B^\xi), \xi = 1, \dots, p\}$ é tal que $A^\xi = B^\xi$ para todo $\xi = 1, \dots, p$, se considerarmos medidas de similaridade $S_U = S_V$, matrizes $G = H$ e parâmetros $\alpha = \beta$ na definição da GEB-FAM, obtemos a GRE-FAM.

Uma propriedade muito importante para um modelo de memória associativa é o armazenamento correto das memórias fundamentais. Lembremos que, dado um conjunto de memórias fundamentais $\mathcal{X} = \{(A^\xi, B^\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$, dizemos que o par $(A^\nu, B^\nu) \in \mathcal{X}$ foi corretamente armazenado em uma memória heteroassociativa se, ao apresentarmos A^ν como padrão de entrada, tal modelo produz B^ν como padrão de saída. Note que, dado A^ν como padrão de entrada, a GEB-FAM produzirá $Y_0 = B^\nu$ se $w_{\nu 0}^A = 1$ e $w_{\xi 0}^A = 0$, para todo $\xi = 1, \dots, p$, $\xi \neq \nu$, em outras palavras, se o vetor de pesos $\mathbf{w}_0 = [w_{10}, \dots, w_{\xi 0}, \dots, w_{p0}]^T$ for igual a ν -ésima coluna da matriz identidade. Ainda, observemos que a operação

$$\sum_{\mu=1}^p g_{\xi\mu} e^{\alpha S_U(A^\mu, X_0)}$$

no cálculo de $w_{\xi 0}^A$, conforme (4.2), corresponde ao produto da ξ -ésima linha da matriz G por $E(X_0)$, em que $E(X)$ é um vetor coluna cuja ξ -ésima componente é dada por $e^{\alpha S_U(A^\xi, X)}$, para $\xi = 1, \dots, p$. Dessa forma, dado A^ν como entrada, podemos obter o vetor

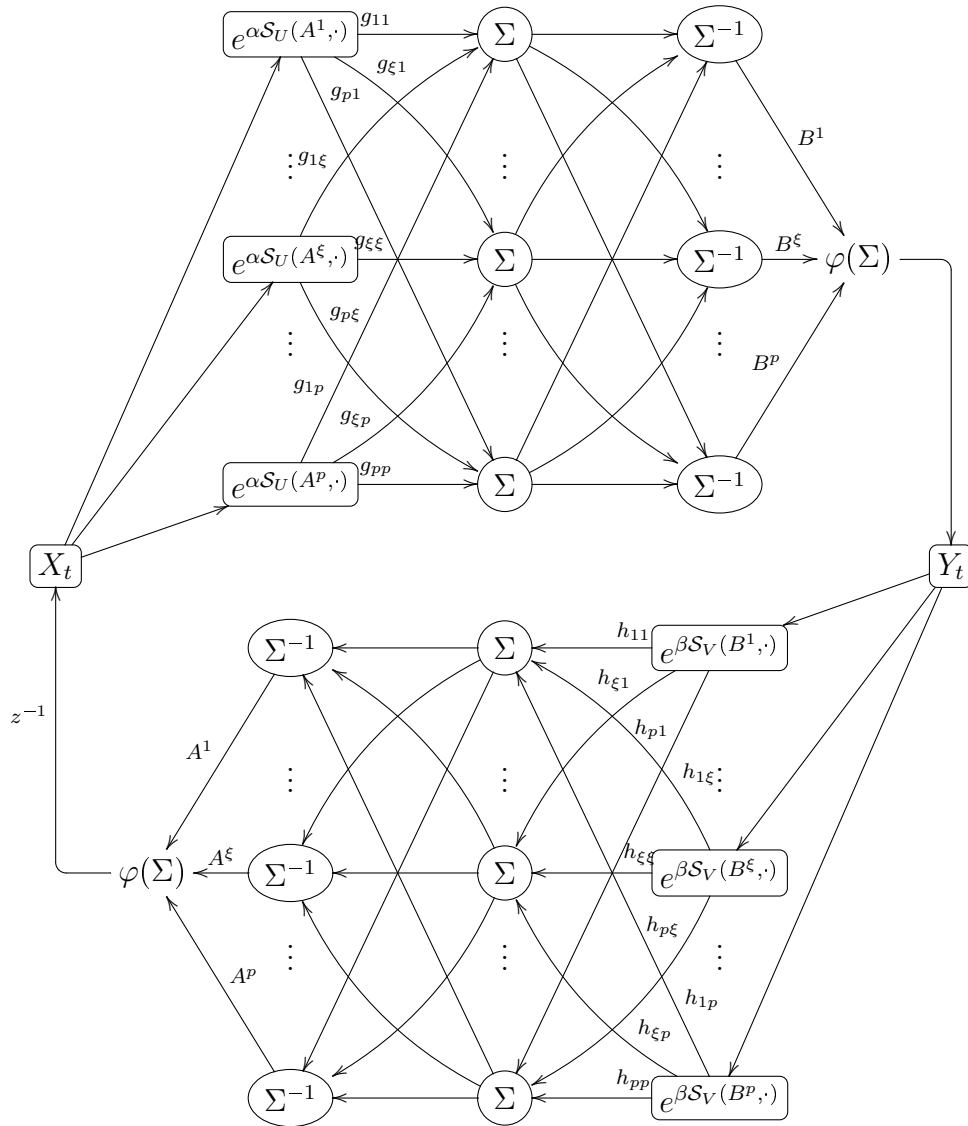


Figura 11 – Diagrama de blocos de uma GEB-FAM.

de pesos \mathbf{w}_0 igual a ν -ésima coluna da matriz identidade, definindo $G = C^{-1}$ em que C é a matriz cuja ξ -ésima coluna é dada pelo vetor $E(A^\xi)$, para $\xi = 1, \dots, p$. Inspirados por essas ideias, o próximo resultado mostra que, se a matriz G for definida adequadamente, a GEB-FAM apresenta capacidade ótima de armazenamento.

Teorema 1. *Seja $\{(A^\xi, B^\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais, com $A^\xi \neq A^\mu$, para todo $\xi, \mu = 1, \dots, p$ com $\xi \neq \mu$. Além disso, sejam $\alpha > 0$ um número real e $\mathcal{S}_U : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ uma medida de similaridade. Consideremos a matriz real $C = (c_{ij})_{p \times p}$ definida por*

$$c_{ij} = e^{\alpha \mathcal{S}_U(A^i, A^j)} \quad (4.4)$$

e suponhamos que C seja inversível. Se $G = C^{-1}$ então, dado o padrão de entrada $X_0 = A^\xi$,

temos que $Y_0 = B^\xi$.

Demonstração. Suponhamos que a matriz C dada por (4.4) seja inversível. Se $G = C^{-1}$, então a seguinte identidade é válida para quaisquer $\xi, \nu \in \{1, \dots, p\}$:

$$\sum_{\mu=1}^p g_{\xi\mu} e^{\alpha S_U(A^\mu, A^\nu)} = G(\xi, :) C(:, \nu) = \delta_{\xi\nu},$$

onde $G(\xi, :)$ denota a ξ -ésima linha da matriz G , $C(:, \nu)$ denota a ν -ésima coluna da matriz C e $\delta_{\xi\nu}$ é a função delta de Kronecker definida por:

$$\delta_{\xi\nu} = \begin{cases} 1, & \xi = \nu, \\ 0, & \xi \neq \nu. \end{cases}$$

Se considerarmos $X_0 = A^\nu$, para algum $\nu \in \{1, \dots, p\}$, então:

$$\begin{aligned} Y_0(v) &= \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\mu=1}^p B^\xi(v) g_{\xi\mu} e^{\alpha S_U(A^\mu, X_0)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} e^{\alpha S_U(A^\mu, X_0)}} \right) = \varphi \left(\frac{\sum_{\xi=1}^p B^\xi(v) \left[\sum_{\mu=1}^p g_{\xi\mu} e^{\alpha S_U(A^\mu, A^\nu)} \right]}{\sum_{\eta=1}^p \left[\sum_{\mu=1}^p g_{\eta\mu} e^{\alpha S_U(A^\mu, A^\nu)} \right]} \right) \\ &= \varphi \left(\frac{\sum_{\xi=1}^p B^\xi(v) \delta_{\xi\nu}}{\sum_{\eta=1}^p \delta_{\eta\nu}} \right) = \varphi(B^\nu(v)) = B^\nu(v), \end{aligned}$$

para todo $v \in V$.

Portanto, se $X_0 = A^\nu$ então $Y_0 = B^\nu$. Em outras palavras, o par (A^ν, B^ν) , $\nu \in \{1, \dots, p\}$, é corretamente armazenado pela GEB-FAM com $G = C^{-1}$. \square

Observação 5. Note que a condição $A^\xi \neq A^\mu$, para todo $\xi, \mu = 1, \dots, p$ com $\xi \neq \mu$ na hipótese do Teorema 1 é essencial. De fato, se $A^\xi = A^\mu$ para $\xi \neq \mu$, teríamos que a ξ -ésima e μ -ésima linhas da matriz C seriam iguais e, portanto, a matriz C não seria inversível.

Observação 6. A matriz H presente na definição de uma GEB-FAM é necessária apenas para obter X_{t+1} a partir de Y_t . Por este motivo, não foi necessário mencionar a matriz H no Teorema 1. No entanto, se $B^\xi \neq B^\mu$, para todo $\xi, \mu = 1, \dots, p$ com $\xi \neq \mu$ podemos considerar $H = D^{-1}$ em que $D = (d_{ij})_{p \times p}$ é definida por $d_{ij} = e^{\beta S_V(B^i, B^j)}$.

Observação 7. Note que a matriz H sugerida na Observação 6 nem sempre é adequada para problemas de classificação. De fato, em um problema de classificação, dado o conjunto das memórias fundamentais $\mathcal{X} = \{(A^\xi, B^\xi), \xi = 1, \dots, p\}$, geralmente B^ξ indica a classe a qual A^ξ pertence. Dessa forma, se houver mais de um representante de alguma das classes no conjunto das memórias fundamentais, a condição $B^\xi \neq B^\mu$, para todo $\xi, \mu = 1, \dots, p$ não é satisfeita e, portanto, a matriz D proposta na Observação 6 não é inversível.

O próximo resultado mostra que o padrão Y_0 produzido em um único passo pela GEB-FAM, quando consideramos \mathcal{S}_U uma medida de similaridade forte, tende para a média aritmética dos B^ξ s associados aos A^ξ s mais similares ao padrão de entrada X_0 quando o parâmetro α tende a infinito. Recordemos que uma medida de similaridade forte é uma função $\mathcal{S}_U : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ simétrica tal que, dados $A, X \in \mathcal{F}(U)$, temos $\mathcal{S}_U(A, X) = 1$ se, e somente se, $A = X$.

Teorema 2. Consideremos $\mathcal{A} = \{(A^1, B^1), \dots, (A^p, B^p)\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais e \mathcal{S}_U uma medida de similaridade forte. Suponhamos que a matriz C definida por (4.4) seja inversível para qualquer $\alpha > 0$. Dado um padrão de entrada X_0 , se $G = C^{-1}$ então

$$\lim_{\alpha \rightarrow \infty} Y_0(v) = \frac{1}{\text{Card}(\Gamma)} \sum_{\gamma \in \Gamma} B^\gamma(v), \quad \forall v \in U, \quad (4.5)$$

onde $Y_0 \in \mathcal{F}(V)$ é o padrão produzido pela GEB-FAM em um único passo a partir de $X_0 \in \mathcal{F}(U)$, $\Gamma = \{\gamma : \mathcal{S}_U(A^\gamma, X_0) \geq \mathcal{S}(A^\xi, X_0), \forall \xi = 1, \dots, p\}$ é o conjunto dos índices dos conjuntos fuzzy A^ξ mais similares a X_0 e $\text{Card}(\Gamma)$ denota a cardinalidade do conjunto Γ .

Demonstração. Seja $\sigma = \max_{\xi=1:p} \{\mathcal{S}_U(A^\xi, X_0)\}$ a maior similaridade entre o conjunto fuzzy de entrada X_0 e as memórias fundamentais A^1, \dots, A^p . A saída Y_0 de um único passo da GEB-FAM satisfaz:

$$Y_0(v) = \varphi \left(\sum_{\xi=1}^p w_{\xi 0}^A B^\xi(v) \right), \quad (4.6)$$

onde $w_{\xi 0}^A$, para $\xi = 1, \dots, p$, é dado por:

$$w_{\xi 0}^A = \frac{\sum_{\mu=1}^p g_{\xi \mu} e^{\alpha \mathcal{S}_U(A^\mu, X_0)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta \mu} e^{\alpha \mathcal{S}_U(A^\mu, X_0)}}. \quad (4.7)$$

Multiplicando o numerador e o denominador de (4.7) por $e^{-\alpha \sigma}$ e reescrevendo as somas, obtemos:

$$w_{\xi 0}^A = \frac{\sum_{\gamma \in \Gamma} g_{\xi \gamma} + \sum_{\mu \notin \Gamma} g_{\xi \mu} e^{\alpha(\mathcal{S}_U(A^\mu, X_0) - \sigma)}}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} g_{\eta \gamma} + \sum_{\eta=1}^p \sum_{\mu \notin \Gamma} g_{\eta \mu} e^{\alpha(\mathcal{S}_U(A^\mu, X_0) - \sigma)}}. \quad (4.8)$$

Agora, a matriz C definida por (4.4) pode ser escrita como $C = e^\alpha D(\alpha)$, em que as entradas de $D(\alpha)$ são $d_{\nu \mu}(\alpha) = e^{\alpha(\mathcal{S}_U(A^\nu, A^\mu) - 1)}$. Além disso, $G = C^{-1} = e^{-\alpha} H(\alpha)$, onde $H(\alpha) = D^{-1}(\alpha)$. Portanto, colocando $e^{-\alpha}$ em evidência, obtemos de (4.8):

$$w_{\xi 0}^A = \frac{\sum_{\gamma \in \Gamma} h_{\xi \gamma}(\alpha) + \sum_{\mu \notin \Gamma} h_{\xi \mu}(\alpha) e^{\alpha(\mathcal{S}_U(A^\mu, X_0) - \sigma)}}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} h_{\eta \gamma}(\alpha) + \sum_{\eta=1}^p \sum_{\mu \notin \Gamma} h_{\eta \mu}(\alpha) e^{\alpha(\mathcal{S}_U(A^\mu, X_0) - \sigma)}}. \quad (4.9)$$

Lembrando que $\mathcal{S}_U(A^\mu, X_0) - \sigma < 0$ para todo $\mu \notin \Gamma$, a segunda soma tanto do numerador quanto no denominador tende a 0 quando α tende a infinito. Além disso, desde que \mathcal{S}_U é uma medida de similaridade forte, $\lim_{\alpha \rightarrow \infty} h_{\nu\mu}(\alpha) = \delta_{\nu\mu}$, onde $h_{\nu\mu}(\alpha)$ é a entrada (ν, μ) de $H(\alpha)$ e $\delta_{\nu\mu}$ denota a função delta de Kronecker.

Portanto,

$$\begin{aligned} \lim_{\alpha \rightarrow \infty} Y_0(v) &= \lim_{\alpha \rightarrow \infty} \varphi \left(\sum_{\xi=1}^p w_{\xi 0}^A B^\xi(v) \right) = \varphi \left(\sum_{\xi=1}^p \left[\lim_{\alpha \rightarrow \infty} \frac{\sum_{\gamma \in \Gamma} h_{\xi\gamma}}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} h_{\eta\gamma}} \right] B^\xi(v) \right) \\ &= \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\gamma \in \Gamma} \delta_{\xi\gamma} B^\xi(v)}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} \delta_{\eta\gamma}} \right) = \frac{\sum_{\gamma \in \Gamma} B^\gamma(v)}{\text{Card}(\Gamma)}, \end{aligned}$$

para todo $v \in V$. Aqui, usamos o fato de que φ é uma função contínua e $\varphi(x) = x$ para todo $x \in [0, 1]$. \square

Corolário 1. Consideremos $\mathcal{A} = \{(A^1, B^1), \dots, (A^p, B^p)\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais e \mathcal{S}_U uma medida de similaridade forte. Suponhamos que a matriz C definida por (4.4) seja inversível para qualquer $\alpha > 0$ e consideremos $G = C^{-1}$. Dado um padrão de entrada X_0 , se existe um único índice ν tal que $\mathcal{S}_U(A^\nu, X_0) > \mathcal{S}(A^\xi, X_0)$, $\forall \xi = 1, \dots, p$, $\xi \neq \nu$, isto é, se existe um vetor A^ν mais similar a X_0 , então

$$\lim_{\alpha \rightarrow \infty} Y_0(v) = B^\nu(v), \quad \forall v \in U, \quad (4.10)$$

onde $Y_0 \in \mathcal{F}(V)$ é o padrão produzido pela GEB-FAM em um único passo a partir de $X_0 \in \mathcal{F}(U)$.

Demonstração. Por hipótese, $\Gamma = \{\gamma : \mathcal{S}_U(A^\gamma, X_0) \geq \mathcal{S}(A^\xi, X_0), \forall \xi = 1, \dots, p\} = \{\nu\}$. Assim, pelo Teorema 2, segue que

$$\lim_{\alpha \rightarrow \infty} Y_0(v) = B^\nu(v), \quad \forall v \in U. \quad (4.11)$$

\square

Observação 8. O Teorema 2 apresenta uma caracterização para o primeiro passo da GEB-FAM quando α tende a infinito. No entanto, para determinar a matriz C sugerida pelo Teorema 1, por exemplo, precisamos calcular $e^{\alpha \mathcal{S}(A^\xi, A^\eta)}$, $\xi, \eta = 1, \dots, p$, valores que tendem a infinito quando α tende a infinito. Assim, na prática, o valor de α é limitado por problemas de overflow. O maior valor de α que pode ser considerado, usando aritmética de ponto flutuante com precisão dupla, está em torno de 700. No próximo capítulo, apresentaremos uma solução para este problema.

Tabela 2 – Valores das similaridades entre o padrão de entrada X_0 e as memórias fundamentais A^ξ , $\xi = 1, \dots, 5$.

$\mathcal{S}_G(A^1, X_0)$	$\mathcal{S}_G(A^2, X_0)$	$\mathcal{S}_G(A^3, X_0)$	$\mathcal{S}_G(A^4, X_0)$	$\mathcal{S}_G(A^5, X_0)$
0,6219	0,3454	0,4132	0,1905	0,2895

Uma vez que as GRE-FAMs podem ser obtidas a partir das GEB-FAMs considerando $B^\xi = A^\xi$, para todo $\xi \in \{1, \dots, p\}$, conforme o Teorema 1, tanto as GRE-FAMs como as GEB-FAMs armazenam corretamente o conjunto das memórias fundamentais quando consideramos a definição adequada da matriz G . No entanto, a dinâmica das GRE-FAMs e GEB-FAMs ainda não está totalmente compreendida. De fato, o próximo exemplo, apresentado em [64], mostra que a GRE-FAM pode produzir uma sequência que não seja convergente se valor do parâmetro α não for suficientemente grande.

Exemplo 3. Consideremos o conjunto de memórias fundamentais $\{A^i, i = 1, \dots, 5\} \subset \mathcal{F}(\{1, 2\})$, formado pelos conjuntos fuzzy:

$$A^1 = \begin{bmatrix} 0,85 \\ 0,2607 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0,5944 \\ 0,0225 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 0,4253 \\ 0,3127 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 0,1615 \\ 0,1788 \end{bmatrix}, \quad A^5 = \begin{bmatrix} 0,4229 \\ 0,0942 \end{bmatrix}$$

e a GRE-FAM obtida utilizando a medida de similaridade de Gregson e a matriz $G = C^{-1}$, em que C é a matriz definida por (4.4). Apresentamos como entrada o vetor

$$X_0 = \begin{bmatrix} 0,9063 \\ 0,8797 \end{bmatrix}$$

e iteramos (3.17) até obtermos $\|X_{t+1} - X_t\| \leq 10^{-6}$ ou $t = 110$.

A Tabela 2 mostra os valores da medida de similaridade entre as memórias fundamentais A^ξ , $\xi = 1, \dots, 5$ e o vetor X_0 . Note que A^1 é a memória fundamental mais similar a X_0 .

A Figura 12, extraída de [64], apresenta as sequências produzidas pela GRE-FAM com os parâmetros $\alpha = 1$, $\alpha = 3$, $\alpha = 5$ e $\alpha = 10$. As memórias fundamentais são representadas pelos símbolos “**x**”. A memória fundamental A_1 (a mais similar a X_0 segundo a medida de similaridade de Gregson) está representada por “**x**” (azul) e as demais memórias fundamentais por “**x**” (vermelho). O padrão inicial X_0 é representado pelo símbolo “•” (azul), X_1 por “•” (verde) e “ X_t , $t \geq 2$, por “•” (preto). Observemos que, no caso em que $\alpha = 10$, a GRE-FAM produziu uma sequência que convergiu para a memória fundamental A^1 , a mais similar ao padrão de entrada X_0 . No entanto, para valores menores de α não houve o mesmo comportamento. Por exemplo, para $\alpha = 1$, a sequência produzida pela GRE-FAM convergiu para a memória fundamental A^2 . Para

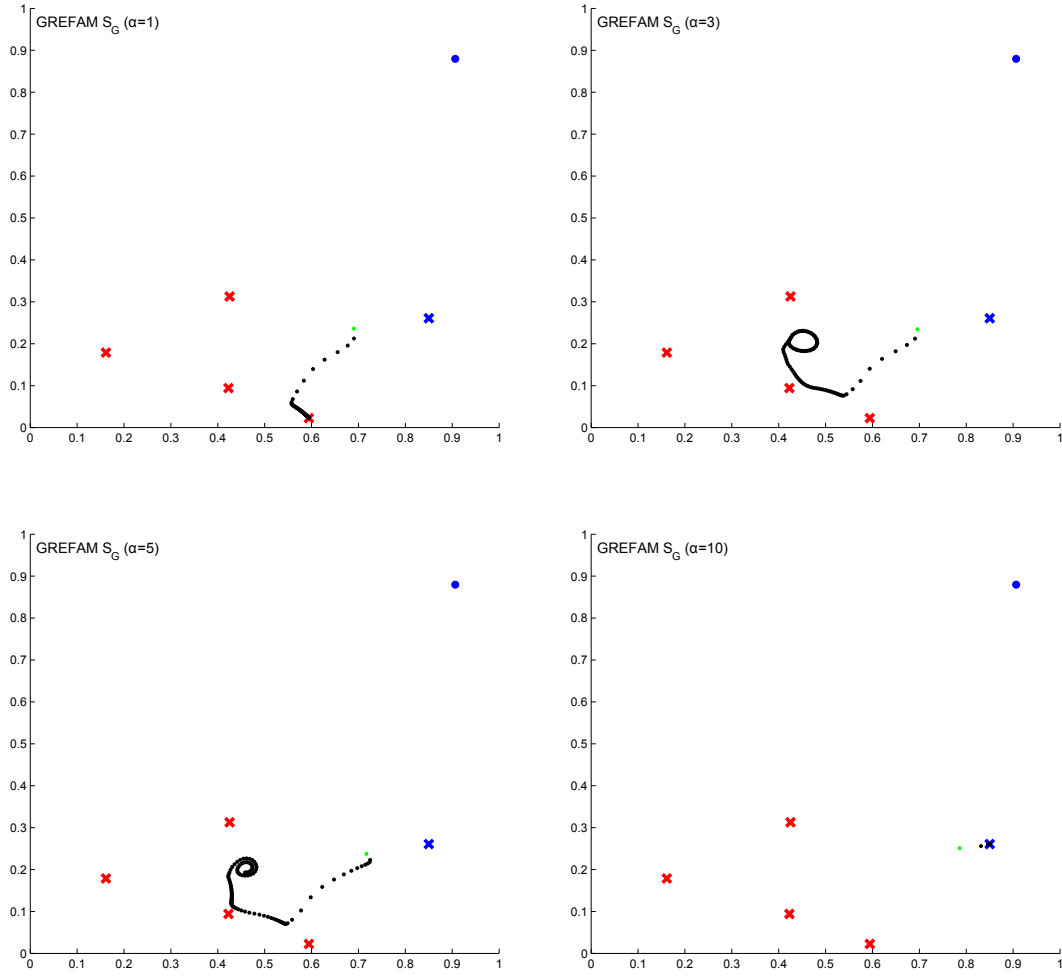


Figura 12 – Sequências produzidas pela GRE-FAM \mathcal{S}_G para $\alpha = 1$, $\alpha = 3$, $\alpha = 5$ e $\alpha = 10$. (Figura extraída de [64]). As memórias fundamentais são representadas pelos símbolos “x”. A memória fundamental A_1 está representada por “x” (azul) e as demais memórias fundamentais por “x” (vermelho). O padrão inicial X_0 é representado pelo símbolo “•” (azul), X_1 por “•” (verde) e “ $X_t, t \geq 2$, por “•” (preto).

$\alpha = 3$, houve um ciclo e portanto a GRE-FAM produziu uma sequência não convergente. Finalmente, com $\alpha = 5$, a GRE-FAM forneceu uma sequência que convergiu para um padrão distinto das memórias fundamentais.

O Exemplo 3 ilustra um caso em que a sequência produzida pela GRE-FAM apresenta diferentes comportamentos conforme o valor do parâmetro α é alterado. Uma vez que a GRE-FAM é um caso particular da GEB-FAM, podemos concluir que tal situação também pode ocorrer com a GEB-FAM. Entretanto, nos experimentos computacionais, valores pequenos de α produziram os melhores resultados. Estes fatos motivaram um estudo mais aprofundado do primeiro passo das GEB-FAMs. De fato, interpretaremos as GEB-FAMs de um único passo utilizando um núcleo *fuzzy*. O modelo obtido, denominado

fuzzy-KAM (do inglês *fuzzy kernel associative memory*), será descrito no próximo capítulo.

5 Memórias Associativas Fuzzy com Núcleo (Fuzzy-KAM)

Neste capítulo introduziremos as memórias associativas *fuzzy* com núcleo (*fuzzy-KAM*, do inglês *fuzzy kernel associative memory*). Este modelo, equivalente a uma GEB-FAM de passo único, utiliza em sua definição uma função a qual nos referiremos como núcleo *fuzzy*, conforme descreveremos a seguir. Ainda, com o objetivo de fornecer uma caracterização do ponto de vista teórico, apresentaremos uma reformulação de teoremas apresentados no capítulo anterior para o caso da *fuzzy-KAM*. Além disso, proporemos o ajuste do parâmetro α da *fuzzy-KAM* com base no conceito de entropia.

5.1 Medidas de Similaridade e Núcleos *Fuzzy*

Métodos de estimação e aprendizagem baseados em núcleos tem se tornado cada vez mais conhecidos e utilizados [22,31,86]. Isto se deve ao fato de que muitos métodos de aprendizado de máquina e estatística contemplam o caso linear, no entanto, muitos problemas reais exigem métodos não lineares [31]. Um núcleo (positivo definido) pode ser interpretado como um produto interno em um espaço de característica de alta dimensão. Neste espaço, o problema pode ser abordado como no caso linear porém, utilizando o núcleo, os cálculos podem ser efetuados com os elementos do espaço original [31].

Como já visto anteriormente, uma medida de similaridade (*fuzzy*) é uma função que associa a um dado par de conjuntos *fuzzy* um número que representa o grau de igualdade desses conjuntos. De acordo com a definição proposta por De Baets e de Meyer em [12], uma medida de similaridade *fuzzy* é uma relação binária *fuzzy* simétrica definida sobre uma família de conjuntos *fuzzy* $\mathcal{F}(U)$, isto é, uma medida de similaridade é uma função simétrica $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$.

Dada uma medida de similaridade \mathcal{S} e um número real $\alpha > 0$, consideremos a função $\kappa : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ definida por:

$$\kappa(A, B) = e^{\alpha(\mathcal{S}(A,B)-1)}. \quad (5.1)$$

Observemos que, uma vez que \mathcal{S} é simétrica, κ também é simétrica. Além disso, como $\mathcal{S}(A, B) \in [0, 1]$ para todo par $(A, B) \in \mathcal{F}(U) \times \mathcal{F}(U)$, segue que $\kappa(A, B) \in [0, 1]$ para todo par $(A, B) \in \mathcal{F}(U) \times \mathcal{F}(U)$. Assim, a função κ satisfaz a definição proposta por De Baets e De Meyer e, portanto, também é uma medida de similaridade. Interpretaremos uma medida de similaridade *fuzzy* como um núcleo *fuzzy*. No que segue, nos referiremos à função κ , definida pela Equação (5.1), como núcleo *fuzzy* (*fuzzy kernel*).

5.2 Definição e Propriedades de uma Fuzzy-KAM

Dada uma medida de similaridade fuzzy, uma memória associativa fuzzy com núcleo (fuzzy-KAM) é definida como segue:

Definição 18 (Fuzzy-KAM). *Consideremos um conjunto de memórias fundamentais $\{(A^\xi, B^\xi) : \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$, $\alpha > 0$ um número real e $G = (g_{ij})_{p \times p}$ uma matriz real de dimensão $p \times p$. Consideremos também uma medida de similaridade $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ e definamos o núcleo fuzzy κ de acordo com (5.1). Uma fuzzy-KAM é uma aplicação $\mathcal{K} : \mathcal{F}(U) \rightarrow \mathcal{F}(V)$ definida pela equação a seguir, onde $X \in \mathcal{F}(U)$ é o conjunto fuzzy de entrada e $Y = \mathcal{K}(X) \in \mathcal{F}(V)$ é a saída produzida a partir de X :*

$$Y(v) = \mathcal{K}(X)(v) = \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\mu=1}^p g_{\xi\mu} \kappa(A^\mu, X) B^\xi(v)}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} \kappa(A^\mu, X)} \right), \quad \forall v \in V, \quad (5.2)$$

em que $\varphi(x) = \max(0, \min(1, x))$ garante que $Y(v) \in [0, 1]$ para todo $v \in V$.

Alternativamente, podemos escrever a saída da fuzzy-KAM como

$$Y(v) = \varphi \left(\sum_{\xi=1}^p w_\xi B^\xi(v) \right), \quad v \in V \quad (5.3)$$

com

$$w_\xi = \frac{\sum_{\mu=1}^p g_{\xi\mu} \kappa(A^\mu, X)}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} \kappa(A^\mu, X)}. \quad (5.4)$$

Observemos que uma fuzzy-KAM autoassociativa é equivalente ao primeiro passo de uma GRE-FAM projetada para armazenar uma família finita de conjuntos fuzzy $\mathcal{A} = \{A^i, i = 1, \dots, p\} \subset \mathcal{F}(U)$ [70], enquanto que a fuzzy-KAM heteroassociativa é equivalente a um único passo da GEB-FAM [65].

Uma fuzzy-KAM pode ser interpretada como uma rede neural artificial com três camadas ocultas. A primeira camada oculta calcula o valor do núcleo fuzzy κ entre a entrada X e os conjuntos fuzzy A^ξ , $\xi = 1, \dots, p$. A segunda camada é composta por neurônios lineares. A terceira camada oculta normaliza a saída fornecida pela camada anterior. Finalmente, utilizando os pesos fornecidos pela terceira camada oculta, a camada de saída produz uma combinação linear dos conjuntos fuzzy B^ξ e aplica a função φ no vetor resultante para determinar a saída Y . A Figura 13 mostra um diagrama de blocos da fuzzy-KAM. No diagrama, Σ^{-1} simboliza a normalização realizada no cálculo dos pesos w_ξ .

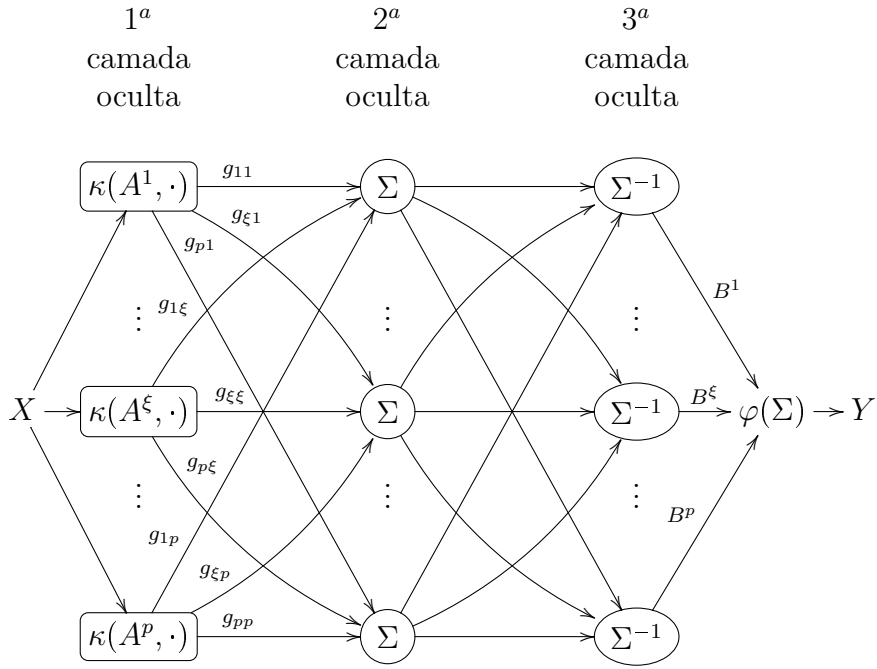


Figura 13 – Diagrama de blocos de uma fuzzy-KAM.

A matriz G desempenha um papel muito importante para a capacidade de armazenamento e tolerância a ruídos da fuzzy-KAM. O próximo resultado, o qual se trata de uma reformulação do Teorema 1 para o caso da fuzzy-KAM, apresenta uma matriz G para a qual todas as memórias fundamentais são armazenadas corretamente pela fuzzy-KAM.

Teorema 3. Consideremos $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{F}(V)$ um conjunto de memórias fundamentais, em que $A^\xi \neq A^\eta$, para $\xi \neq \eta$. Seja $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ uma medida de similaridade, $\alpha > 0$ um número real e $\kappa : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ o núcleo fuzzy definido por (5.1). Se a matriz $K = (k_{ij}) \in \mathbb{R}^{p \times p}$ cujas entradas são definidas por

$$k_{ij} = \kappa(A^i, A^j), \quad \forall i, j = 1, \dots, p, \quad (5.5)$$

for inversível, então a fuzzy-KAM obtida considerando $G = K^{-1}$ satisfaz a identidade $\mathcal{K}(A^\xi) = B^\xi$ para todo $\xi = 1, \dots, p$.

Demonstração. Suponhamos que a matriz K dada por (5.5) seja inversível. Se $G = K^{-1}$, então, a seguinte identidade é válida para todos os índices $\xi, \nu \in \{1, \dots, p\}$:

$$\sum_{\mu=1}^p g_{\xi\mu} \kappa(A^\mu, A^\nu) = \sum_{\mu=1}^p g_{\xi\mu} k_{\mu\nu} = G(\xi, \cdot) K(\cdot, \nu) = \delta_{\xi\nu},$$

onde $G(\xi, \cdot)$ denota a ξ -ésima linha da matriz G , $K(\cdot, \nu)$ denota ν -ésima coluna da matriz K e $\delta_{\xi\nu}$ é a função delta de Kronecker definida por:

$$\delta_{\xi\nu} = \begin{cases} 1, & \xi = \nu, \\ 0, & \xi \neq \nu. \end{cases}$$

Dado um conjunto de entrada $X = A^\nu$, $\nu \in \{1, \dots, p\}$, a *fuzzy-KAM* produz:

$$\begin{aligned} Y = \mathcal{K}(A^\nu) &= \varphi \left(\frac{\sum_{\xi=1}^p \sum_{\mu=1}^p B^\xi g_{\xi\mu} \kappa(A^\mu, A^\nu)}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} \kappa(A^\mu, A^\nu)} \right) = \varphi \left(\frac{\sum_{\xi=1}^p B^\xi \left[\sum_{\mu=1}^p g_{\xi\mu} \kappa(A^\mu, A^\nu) \right]}{\sum_{\eta=1}^p \left[\sum_{\mu=1}^p g_{\eta\mu} \kappa(A^\mu, A^\nu) \right]} \right) \\ &= \varphi \left(\frac{\sum_{\xi=1}^p B^\xi \delta_{\xi\nu}}{\sum_{\eta=1}^p \delta_{\eta\nu}} \right) = \varphi(B^\nu) = B^\nu. \end{aligned}$$

Portanto, se $G = K^{-1}$ então $\mathcal{K}(A^\nu) = B^\nu$, para todo $\nu = 1, \dots, p$. \square

Observação 9. Conforme mencionamos na Observação 8, na prática, podemos ter problemas de overflow ao definir a matriz C sugerida pelo Teorema 1, caso α seja muito grande. No entanto, como a imagem da função κ está contida no intervalo $[0, 1]$, este problema não ocorre na definição da matriz K dada por (5.5).

O parâmetro α também tem um papel importante na capacidade de generalização da *fuzzy-KAM*, uma vez que, de certa forma, o parâmetro α controla como cada memória fundamental contribui para produzir a saída de uma *fuzzy-KAM*. Em outras palavras, aumentar o valor de α significa enfatizar a contribuição das memórias fundamentais mais similares ao conjunto de entrada. O próximo teorema afirma que, quando α tende a infinito, a saída $\mathcal{K}(X)$ converge pontualmente para a média aritmética das respostas desejadas B^ξ associadas aos A^ξ mais similares ao padrão de entrada X .

Teorema 4. Seja $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \dots, p\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ a família das memórias fundamentais. Consideremos \mathcal{S} uma medida de similaridade forte e definamos o núcleo fuzzy κ conforme (5.1). Suponhamos que a matriz K , dada por (5.5), seja inversível para qualquer $\alpha > 0$ e consideremos $G(\alpha) = K^{-1}(\alpha)$. Dado um conjunto fuzzy $X \in \mathcal{F}(U)$, seja $\Gamma \subseteq \{1, \dots, p\}$ o conjunto dos índices dos estímulos A^ξ mais similares à entrada X em termos de \mathcal{S} , isto é:

$$\Gamma = \{\gamma : \mathcal{S}(A^\gamma, X) \geq \mathcal{S}(A^\xi, X), \forall \xi = 1, \dots, p\}. \quad (5.6)$$

Então,

$$\lim_{\alpha \rightarrow \infty} Y(v) = \frac{1}{\text{Card}(\Gamma)} \sum_{\gamma \in \Gamma} B^\gamma(v), \quad \forall v \in V, \quad (5.7)$$

onde $Y = \mathcal{K}(X)$ é a saída de uma fuzzy-KAM. Além disso, os pesos w_ξ dados por (5.4) satisfazem a seguinte equação para todo $\xi = 1, \dots, p$:

$$\lim_{\alpha \rightarrow \infty} w_\xi = \begin{cases} \frac{1}{\text{Card}(\Gamma)}, & \xi \in \Gamma, \\ 0, & \text{caso contrário.} \end{cases} \quad (5.8)$$

Apesar do Teorema 4 ser, basicamente, uma reformulação do Teorema 2 para as fuzzy-KAMs, incluímos abaixo sua demonstração para leitores que optarem por ler apenas esse capítulo.

Demonstração. Seja $\sigma = \max_{\xi=1:p} \{\mathcal{S}(A^\xi, X)\}$ o maior valor da similaridade entre X e as memórias fundamentais A^1, \dots, A^p . A fuzzy-KAM produz a partir de X um conjunto fuzzy $Y = \mathcal{K}(X)$ dado por:

$$Y(v) = \varphi \left(\sum_{\xi=1}^p w_\xi B^\xi(v) \right), \quad (5.9)$$

onde w_ξ , definido por (5.4), é dado por:

$$w_\xi = \frac{\sum_{\mu=1}^p g_{\xi\mu} e^{\alpha(\mathcal{S}(A^\mu, X) - 1)}}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} e^{\alpha(\mathcal{S}(A^\mu, X) - 1)}}, \quad \forall \xi = 1, \dots, p. \quad (5.10)$$

Multiplicando o numerador e o denominador de (5.10) por $e^{\alpha(1-\sigma)}$ e reorganizando as somas, obtemos:

$$w_\xi = \frac{\sum_{\gamma \in \Gamma} g_{\xi\gamma} + \sum_{\mu \notin \Gamma} g_{\xi\mu} e^{\alpha(\mathcal{S}(A^\mu, X) - \sigma)}}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} g_{\eta\gamma} + \sum_{\eta=1}^p \sum_{\mu \notin \Gamma} g_{\eta\mu} e^{\alpha(\mathcal{S}(A^\mu, X) - \sigma)}}. \quad (5.11)$$

Desde que $\mathcal{S}(A^\mu, X) - \sigma < 0$ para todo $\mu \notin \Gamma$, a segunda soma tanto no numerador como no denominador tende a 0 quando α tende a infinito. Além disso, por hipótese, \mathcal{S} é uma medida de similaridade forte, então $\lim_{\alpha \rightarrow \infty} g_{\nu\mu}(\alpha) = \delta_{\nu\mu}$, onde $g_{\nu\mu}(\alpha)$ é a (ν, μ) entrada de $G(\alpha)$ e $\delta_{\nu\mu}$ denota o delta de Kronecker. Portanto,

$$\lim_{\alpha \rightarrow \infty} w_\xi = \frac{\sum_{\gamma \in \Gamma} \delta_{\xi\gamma}}{\sum_{\eta=1}^p \sum_{\gamma \in \Gamma} \delta_{\eta\gamma}} = \begin{cases} \frac{1}{\text{Card}(\Gamma)}, & \xi \in \Gamma, \\ 0, & \text{caso contrário.} \end{cases} \quad (5.12)$$

Como φ é contínua e $\varphi(x) = x$ para todo $x \in [0, 1]$, concluímos que

$$\lim_{\alpha \rightarrow \infty} Y(v) = \lim_{\alpha \rightarrow \infty} \varphi \left(\sum_{\xi=1}^p w_{\xi} B^{\xi}(v) \right) = \varphi \left(\sum_{\xi=1}^p \left[\lim_{\alpha \rightarrow \infty} w_{\xi} \right] B^{\xi}(v) \right) \quad (5.13)$$

$$= \varphi \left(\sum_{\xi \in \Gamma} \frac{1}{\text{Card}(\Gamma)} B^{\xi}(v) \right) = \frac{1}{\text{Card}(\Gamma)} \sum_{\gamma \in \Gamma} B^{\gamma}(v). \quad (5.14)$$

para todo $v \in V$. \square

A importância do parâmetro α para o modelo motivou um estudo mais aprofundado de uma maneira de definir seu valor. Inicialmente, buscamos definir α para que a matriz K dada por (5.5) fosse definida positiva, condição que, além de garantir a existência da matriz $G = K^{-1}$, contribuiria para a redução do custo computacional. Se considerarmos uma medida de similaridade forte, o próximo resultado nos fornece uma condição suficiente para os valores de α a partir do qual a matriz K definida por (5.5) é definida positiva.

Teorema 5. *Sejam $\mathcal{A} = \{(A^i, B^i) | i = 1, \dots, p\} \subset \mathcal{F}(U)$ o conjunto das memórias fundamentais, $\mathcal{S} : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ uma medida de similaridade forte, $\alpha > 0$ um número real e $K = (k_{ij})$ a matriz definida por:*

$$k_{ij} = e^{\alpha(\mathcal{S}(A^i, A^j) - 1)}, \quad \forall i, j = 1, \dots, p. \quad (5.15)$$

Se $\alpha > \frac{\ln(p-1)}{1 - S_m}$ então a matriz K é definida positiva, onde p denota o número de memórias fundamentais e $S_m = \max_{i \neq j} \{\mathcal{S}(A^i, A^j)\}$.

Demonstração. Primeiramente, observemos que uma matriz simétrica é definida positiva se, e somente se, todos os seus autovalores são positivos [23]. Dessa forma, mostraremos que se $\alpha > \frac{-\ln(p-1)}{S_m - 1}$, todos os autovalores de K são positivos.

De fato, seja λ um autovalor qualquer de K . Aplicando o teorema dos discos de Gershgorin [52] à matriz K , concluímos que existe $i \in \{1, \dots, p\}$ tal que:

$$|1 - \lambda| \leq \sum_{\substack{j=1 \\ j \neq i}}^p e^{\alpha(\mathcal{S}(A^i, A^j) - 1)}. \quad (5.16)$$

De forma equivalente, temos

$$-\sum_{\substack{j=1 \\ j \neq i}}^p e^{\alpha(\mathcal{S}(A^i, A^j) - 1)} \leq 1 - \lambda \leq \sum_{\substack{j=1 \\ j \neq i}}^p e^{\alpha(\mathcal{S}(A^i, A^j) - 1)}. \quad (5.17)$$

Da segunda desigualdade acima, desde que $S_m = \max_{i, j, i \neq j} \{\mathcal{S}(A^i, A^j)\} \geq \mathcal{S}(A^i, A^j)$, para quaisquer $i, j \in \{1, \dots, p\}$, $i \neq j$, segue que:

$$1 - \lambda \leq \sum_{\substack{j=1 \\ j \neq i}}^p e^{\alpha(\mathcal{S}(A^i, A^j) - 1)} \leq \sum_{\substack{j=1 \\ j \neq i}}^p e^{\alpha(S_m - 1)} = (p-1)e^{\alpha(S_m - 1)}. \quad (5.18)$$

Assim, temos:

$$1 - (p - 1)e^{\alpha(S_m - 1)} \leq \lambda. \quad (5.19)$$

Por hipótese, temos que $\alpha > \frac{-\ln(p - 1)}{S_m - 1}$. Assim, obtemos $e^{\alpha(S_m - 1)} < (p - 1)^{-1}$, o que implica que $(p - 1)e^{\alpha(S_m - 1)} < 1$. Portanto, $1 - (p - 1)e^{\alpha(S_m - 1)} > 0$.

$$\text{Logo, } 0 < 1 - (p - 1)e^{\alpha(S_m - 1)} \leq \lambda.$$

□

Lembrando que toda matriz definida positiva é inversível [23], temos que se $\alpha > \frac{-\ln(p - 1)}{S_m - 1}$, então a matriz $G = K^{-1}$ está bem definida. Além disso, sendo K simétrica e definida positiva, ela admite fatoração de Cholesky, isto é, K pode ser escrita na forma $K = R^T R$ em que R é uma matriz triangular superior e R^T denota sua transposta [23]. Neste caso, não é necessário calcular a matriz G explicitamente. De fato, se a matriz K é definida positiva, em vez de utilizar $G = K^{-1}$, podemos resolver dois sistemas lineares utilizando o fator de Cholesky R de K . Dessa forma, é possível reduzir o custo computacional da *fuzzy*-KAM principalmente para os casos em que K apresenta dimensão grande, isto é, nos casos em que estamos considerando muitas memórias fundamentais.

No entanto, os testes preliminares mostraram que esta estimativa não fornece valores adequados para α . De fato, em geral, os melhores resultados computacionais foram obtidos para α em torno de 1, valor muito menor do que o obtido com a estimativa acima. Além disso, para valores demasiadamente grandes de α , a memória fundamental mais similar ao padrão de entrada é fortemente enfatizada, de forma que as outras memórias fundamentais praticamente não são consideradas para a produção do padrão de saída, o que pode prejudicar o desempenho do modelo em certas aplicações como classificação e reconhecimento de faces. Assim, buscamos uma outra estratégia para ajustar tal parâmetro.

5.3 Entropia e Ajuste do Parâmetro α

Vamos analisar como ocorre a dependência do núcleo *fuzzy* κ em relação ao parâmetro α quando consideramos \mathcal{S} uma medida de similaridade forte. Primeiramente, observemos que κ se aproxima da igualdade estrita de conjuntos *fuzzy* quando o valor de α aumenta, isto é, quando $\alpha \rightarrow \infty$, temos $\kappa(A, B) = 1$ se $A = B$ e $\kappa(A, B) \rightarrow 0$ caso contrário. Por outro lado, κ é incapaz de discriminar conjuntos *fuzzy* para valores suficientemente pequenos de $\alpha > 0$. De fato, quando $\alpha \rightarrow 0$, temos que $\kappa(A, B) \rightarrow 1$, para todo $A, B \in \mathcal{F}(U)$. Isto nos mostra que o valor do parâmetro α controla a capacidade do núcleo *fuzzy* κ de distinguir entre conjuntos *fuzzy* [67]. Propomos uma forma de estimar o parâmetro α de uma *fuzzy*-KAM utilizando aprendizagem baseada em teoria da informação [58]. A ideia básica é maximizar a capacidade do núcleo *fuzzy* κ em discriminar dois estímulos diferentes. Para isto, utilizaremos o conceito de entropia.

O conceito de entropia foi introduzido por Shannon em 1948 [62] e representa uma medida quantitativa da incerteza e informação de um sistema probabilístico [39, 56]. A entropia de um sistema de n estados é definida pela seguinte equação, onde p_i denota a probabilidade de ocorrência do i -ésimo estado [56]:

$$E = \sum_{i=1}^n p_i \log(1/p_i). \quad (5.20)$$

Inspirados por (5.20), dado um conjunto de memórias fundamentais $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \dots, p\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ e uma medida de similaridade forte \mathcal{S} , definimos a entropia de uma fuzzy-KAM \mathcal{K} por meio da equação:

$$E_{\mathcal{K}}(\alpha) = \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \ln \left(1/\kappa(A^i, A^j) \right) \quad (5.21)$$

$$= \sum_{i=1}^p \sum_{j=1}^p -e^{\alpha(\mathcal{S}(A^i, A^j)-1)} \ln \left(e^{\alpha(\mathcal{S}(A^i, A^j)-1)} \right) \quad (5.22)$$

$$= -\sum_{i=1}^p \sum_{j=1}^p \alpha(\mathcal{S}(A^i, A^j) - 1) e^{\alpha(\mathcal{S}(A^i, A^j)-1)}. \quad (5.23)$$

Observemos que a entropia $E_{\mathcal{K}}$ de uma fuzzy KAM é uma função do parâmetro α e tende a zero se $\alpha \rightarrow 0$ ou $\alpha \rightarrow \infty$. Intuitivamente, podemos dizer que $E_{\mathcal{K}}$ quantifica a capacidade do núcleo fuzzy κ discriminar entre os conjuntos fuzzy A^i e A^j , em que $i \neq j$, como função de α . Dessa forma, com o objetivo de melhorar a tolerância a ruído da fuzzy-KAM, sugerimos a escolha do parâmetro α^* que maximiza (5.21). Formalmente, propomos a definição:

$$\alpha^* = \operatorname{argmax}_{\alpha > 0} E_{\mathcal{K}}(\alpha). \quad (5.24)$$

Uma vez que sugerimos a escolha do parâmetro α como $\operatorname{argmax}_{\alpha > 0} E_{\mathcal{K}}(\alpha)$ é importante mostrar que tal valor está bem definido. O próximo teorema mostra que, se considerarmos uma medida de similaridade forte \mathcal{S} , existe $\alpha^* = \operatorname{argmax}_{\alpha > 0} E_{\mathcal{K}}(\alpha)$.

Teorema 6. *Sejam $\mathcal{A} = \{(A^\xi, B^\xi) : \xi = 1, \dots, p\} \subseteq \mathcal{F}(U) \times \mathcal{F}(V)$ um conjunto de memórias fundamentais e \mathcal{S} uma medida de similaridade forte. A entropia de uma fuzzy-KAM definida por $E_{\mathcal{K}}(\alpha) = -\sum_{i=1}^p \sum_{j=1}^p \alpha(\mathcal{S}(A^i, A^j) - 1) e^{\alpha(\mathcal{S}(A^i, A^j)-1)}$ admite um máximo global $\alpha^* = \operatorname{argmax}_{\alpha > 0} E_{\mathcal{K}}(\alpha)$ em $(0, \infty)$.*

Demonstração. Observemos que a entropia da fuzzy-KAM, dada por

$$E_{\mathcal{K}}(\alpha) = -\sum_{i=1}^p \sum_{j=1}^p \alpha(\mathcal{S}(A^i, A^j) - 1) e^{\alpha(\mathcal{S}(A^i, A^j)-1)},$$

é uma função contínua, diferenciável e sua derivada E'_K é dada por:

$$E'_K(\alpha) = \sum_{i=1}^p \sum_{j=1}^p -(\mathcal{S}(A^i, A^j) - 1)e^{\alpha(\mathcal{S}(A^i, A^j) - 1)} - \alpha(\mathcal{S}(A^i, A^j) - 1)^2 e^{\alpha(\mathcal{S}(A^i, A^j) - 1)}. \quad (5.25)$$

Equivalentemente, temos que

$$E'_K(\alpha) = \sum_{\substack{i,j=1 \\ i \neq j}}^p \underbrace{e^{\alpha(\mathcal{S}(A^i, A^j) - 1)}(1 - \mathcal{S}(A^i, A^j))}_{\geq 0} \underbrace{[1 - \alpha(1 - \mathcal{S}(A^i, A^j))]}_{(*)}. \quad (5.26)$$

Definamos $s_{\min} = \min_{\substack{i,j \\ i \neq j}} \mathcal{S}(A^i, A^j)$ e $s_{\max} = \max_{\substack{i,j \\ i \neq j}} \mathcal{S}(A^i, A^j)$.

Por um lado, se $\alpha < \frac{1}{1 - s_{\min}}$, então:

$$1 - \alpha(1 - \mathcal{S}(A^i, A^j)) > 1 - \frac{1 - \mathcal{S}(A^i, A^j)}{1 - s_{\min}} \geq 0, \quad (5.27)$$

ou seja, o termo $(*)$ em (5.26) é positivo e, consequentemente $E'_K > 0$.

Por outro lado, se $\alpha > \frac{1}{1 - s_{\max}}$, então:

$$1 - \alpha(1 - \mathcal{S}(A^i, A^j)) < 1 - \frac{1 - \mathcal{S}(A^i, A^j)}{1 - s_{\max}} \leq 0. \quad (5.28)$$

Portanto, o termo $(*)$ em (5.26) é negativo e, consequentemente $E'_K < 0$.

Assim, pelo Teorema do Valor Intermediário [25], existe α tal que $E'_K(\alpha) = 0$. Além disso, temos que $\lim_{\alpha \rightarrow 0} E_K(\alpha) = \lim_{\alpha \rightarrow \infty} E_K(\alpha) = 0$. Logo, podemos concluir que existe $\alpha^* = \operatorname{argmax}_{\alpha > 0} E_K(\alpha)$.

□

Observação 10. Em (5.21) consideramos o logaritmo na base e . A escolha desta base simplifica os cálculos, uma vez que o núcleo fuzzy κ é definido por uma função exponencial. No entanto, a escolha de outra base para o logaritmo não altera o valor para o qual a função entropia atinge seu máximo, isto é, não altera o valor do parâmetro α . De fato, se considerássemos em (5.21) o logaritmo em uma base b , \log_b , teríamos:

$$\begin{aligned} \tilde{E}(\alpha) &= \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \log_b \left(\frac{1}{\kappa(A^i, A^j)} \right) = \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \frac{\ln \left(\frac{1}{\kappa(A^i, A^j)} \right)}{\ln b} \\ &= \frac{1}{\ln b} \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \ln \left(\frac{1}{\kappa(A^i, A^j)} \right) = \frac{1}{\ln b} E_K(\alpha). \end{aligned}$$

Logo, teríamos $\operatorname{argmax}_{\alpha > 0} \tilde{E}(\alpha) = \operatorname{argmax}_{\alpha > 0} E_K(\alpha)$.

A definição da função entropia da *fuzzy*-KAM, como visto anteriormente, foi inspirada na definição de entropia (5.20) proposta por Shannon, a qual baseia-se na teoria clássica de probabilidade. De fato, em (5.20), p_i representa a probabilidade de ocorrência do i -ésimo estado de um sistema com n estados possíveis. Assim, temos que $0 \leq p_i \leq 1$, para todo $i = 1, \dots, n$ e $\sum_{i=1}^n p_i = 1$. Na definição da entropia da *fuzzy*-KAM dada por (5.21), temos que $0 \leq \kappa(A^i, A^j) \leq 1$, para todo $i, j = 1, \dots, p$, no entanto a propriedade $\sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) = 1$ não é satisfeita. Uma alternativa para que tal propriedade seja cumprida seria considerar uma normalização do núcleo *fuzzy* κ e definir a entropia da *fuzzy*-KAM por meio da seguinte equação:

$$\overline{E}(\alpha) = \sum_{i=1}^p \sum_{j=1}^p \frac{\kappa(A^i, A^j)}{c} \ln \left(\frac{c}{\kappa(A^i, A^j)} \right), \quad (5.29)$$

onde c é a constante dada por $c = \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j)$. Assim, teríamos:

$$\begin{aligned} \overline{E}(\alpha) &= \sum_{i=1}^p \sum_{j=1}^p \frac{\kappa(A^i, A^j)}{c} \ln \left(\frac{c}{\kappa(A^i, A^j)} \right) \\ &= \frac{1}{c} \left[\sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) (\ln(c) - \ln(\kappa(A^i, A^j))) \right] \\ &= \frac{1}{c} \ln(c) \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) - \frac{1}{c} \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \ln(\kappa(A^i, A^j)) \\ &= \ln(c) + \frac{1}{c} \sum_{i=1}^p \sum_{j=1}^p \kappa(A^i, A^j) \ln \left(\frac{1}{\kappa(A^i, A^j)} \right) = \ln(c) + \frac{1}{c} E_{\mathcal{K}(\alpha)}. \end{aligned}$$

Isto mostra que a função \overline{E} corresponde a uma versão reescalada e transladada da função entropia $E_{\mathcal{K}}$ definida anteriormente. Por este motivo, teríamos $\operatorname{argmax}_{\alpha > 0} \overline{E}(\alpha) = \operatorname{argmax}_{\alpha > 0} E_{\mathcal{K}}(\alpha)$. Em outras palavras, esta mudança na definição da entropia da *fuzzy*-KAM proposta inicialmente não alteraria o valor do parâmetro α .

Em [14] De Luca e Termini propuseram a seguinte definição de entropia de um conjunto *fuzzy* $X = [x_1, \dots, x_n]$:

$$H(X) = -k \sum_{i=1}^n x_i \log(x_i) + (1 - x_i) \log(1 - x_i). \quad (5.30)$$

Conforme descrito em [14], esta definição de entropia *fuzzy* não leva em conta conceitos probabilísticos. Desse ponto de vista, a entropia *fuzzy* seria mais apropriada em nosso contexto. No entanto, nos experimentos preliminares realizados, não observamos melhora significativa nos resultados obtidos usando a entropia *fuzzy* se comparados aos obtidos usando a entropia de Shannon. Além disso, a entropia de Shannon possui menor

custo computacional e mostrou ser mais robusta. Portanto, optamos por definir a entropia da *fuzzy*-KAM com base na entropia de Shannon.

Observemos que a escolha de α usando a entropia pode não garantir que a matriz K definida por (5.5) seja inversível. De fato, o Teorema 5 garante que a matriz K é inversível se considerarmos valores suficientemente grandes do parâmetro α porém, não temos garantia teórica de que K seja inversível para qualquer valor de α . Caso a matriz K obtida para um determinado valor de α não seja inversível, podemos definir a matriz G presente na definição da *fuzzy*-KAM por $G = K^\dagger$, em que K^\dagger denota a pseudo-inversa de K . Na prática, sempre obtivemos matrizes K inversíveis nos experimentos computacionais e, portanto, consideramos $G = K^{-1}$, conforme sugerido pelo Teorema 3. No entanto, o cálculo da pseudo-inversa de K , ou mesmo da inversa ou da fatoração de Cholesky de K , pode ser demasiadamente custoso se o número de memórias fundamentais p for grande. Para contornar esse problema, apresentaremos uma formulação matricial para a *fuzzy*-KAM que permite definir uma matriz G esparsa em blocos.

5.4 Formulação Matricial da *Fuzzy*-KAM

A *fuzzy*-KAM também pode ser descrita por meio de uma formulação matricial, conforme apresentaremos a seguir.

Consideremos universos de discurso finitos $U = \{u_1, \dots, u_n\}$ e $V = \{v_1, \dots, v_m\}$. Sejam $\{(A^\xi, B^\xi), \xi = 1, \dots, p\} \in \mathcal{F}(U) \times \mathcal{F}(V)$ o conjunto das memórias fundamentais, G uma matriz de dimensão $p \times p$ e $\kappa : \mathcal{F}(U) \times \mathcal{F}(U) \rightarrow [0, 1]$ o núcleo *fuzzy* dado por $\kappa(A, X) = e^{\alpha(\mathcal{S}(A, X) - 1)}$, em que \mathcal{S} é uma medida de similaridade e α um número real positivo.

Consideremos também as seguintes funções vetoriais: $E : \mathcal{F}(U) \rightarrow [0, 1]^p$ que transforma $X \in \mathcal{F}(U)$ no vetor coluna

$$E(X) = \begin{bmatrix} \kappa(A^1, X) \\ \kappa(A^2, X) \\ \vdots \\ \kappa(A^p, X) \end{bmatrix},$$

a função $F : \mathbb{R}^p \rightarrow \mathbb{R}^p$ que associa o vetor $X = [x_1, \dots, x_p]^T \in \mathbb{R}^p$ ao vetor normalizado

$$F(X) = \begin{bmatrix} x_1 / \sum_{i=1}^p x_i \\ x_2 / \sum_{i=1}^p x_i \\ \vdots \\ x_p / \sum_{i=1}^p x_i \end{bmatrix}$$

e a função $\phi : \mathbb{R}^m \rightarrow \mathbb{R}^m$ definida por

$$\phi(Y) = \begin{bmatrix} \varphi(y_1) \\ \varphi(y_2) \\ \vdots \\ \varphi(y_m) \end{bmatrix},$$

em que $\varphi(y) = \max(0, \min(1, y))$.

Observemos que dado um conjunto *fuzzy* de entrada X , os pesos w_ξ , $\xi = 1, \dots, p$ definidos por (5.4), podem ser calculados fazendo $F(GE(X))$. Especificamente, temos que $\mathbf{w} = F(GE(X))$, onde \mathbf{w} denota o vetor coluna cuja ξ -ésima componente é w_ξ , $\xi = 1, \dots, p$. Assim, dado o conjunto de entrada $X \in \mathcal{F}(U)$, a saída da *fuzzy*-KAM é dada por

$$Y = \phi(BF(GE(X))), \quad (5.31)$$

onde B denota a matriz de dimensão $m \times p$ cuja i -ésima coluna é o vetor $B^i \in [0, 1]^m$, $i = 1 \dots, p$, do conjunto das memórias fundamentais.

Com o objetivo de deixar mais claro os processos de armazenamento e recordação de uma *fuzzy*-KAM, apresentaremos exemplos simples para explicitar cada um dos passos envolvidos. Primeiramente faremos um exemplo considerando a abordagem autoassociativa e, em seguida, um exemplo para ilustrar a abordagem heteroassociativa. Retomaremos estes exemplos no Capítulo 6.

Exemplo 4. Consideremos o conjunto de memórias fundamentais $\mathcal{A} = \{A^1, A^2, \dots, A^8\} \subset [0, 1]^2$ formado pelos conjuntos fuzzy

$$\begin{aligned} A^1 &= \begin{bmatrix} 0, 1 \\ 0, 5 \end{bmatrix}, & A^2 &= \begin{bmatrix} 0, 2 \\ 0, 4 \end{bmatrix}, & A^3 &= \begin{bmatrix} 0, 2 \\ 0, 5 \end{bmatrix}, & A^4 &= \begin{bmatrix} 0, 2 \\ 0, 6 \end{bmatrix}, \\ A^5 &= \begin{bmatrix} 0, 4 \\ 0, 2 \end{bmatrix}, & A^6 &= \begin{bmatrix} 0, 4 \\ 0, 8 \end{bmatrix}, & A^7 &= \begin{bmatrix} 0, 6 \\ 0, 4 \end{bmatrix}, & A^8 &= \begin{bmatrix} 0, 6 \\ 0, 6 \end{bmatrix}, \end{aligned}$$

a medida de similaridade baseada na norma-1, \mathcal{S}_H , definida por (2.11). Definiremos uma *fuzzy*-KAM autoassociativa e, em seguida, dado um conjunto fuzzy de entrada, descreveremos a etapa de recordação.

O primeiro passo é determinar a matriz $S = (s_{ij})$, em que $s_{ij} = \mathcal{S}_H(A^i, A^j)$:

$$S = \begin{pmatrix} 1 & 0,9 & 0,95 & 0,9 & 0,7 & 0,7 & 0,7 & 0,7 \\ 0,9 & 1 & 0,95 & 0,9 & 0,8 & 0,7 & 0,8 & 0,7 \\ 0,95 & 0,95 & 1 & 0,95 & 0,75 & 0,75 & 0,75 & 0,75 \\ 0,9 & 0,9 & 0,95 & 1 & 0,7 & 0,8 & 0,7 & 0,8 \\ 0,7 & 0,8 & 0,75 & 0,7 & 1 & 0,7 & 0,8 & 0,7 \\ 0,7 & 0,7 & 0,75 & 0,8 & 0,7 & 1 & 0,7 & 0,8 \\ 0,7 & 0,8 & 0,75 & 0,7 & 0,8 & 0,7 & 1 & 0,9 \\ 0,7 & 0,7 & 0,75 & 0,8 & 0,7 & 0,8 & 0,9 & 1 \end{pmatrix}. \quad (5.32)$$

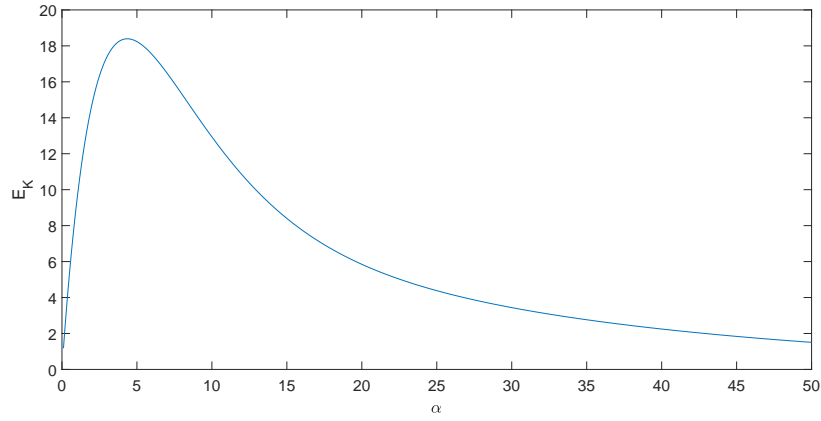


Figura 14 – Entropia da fuzzy-KAM em função do parâmetro α nas condições do Exemplo 4.

Em seguida, calculamos o valor do parâmetro α que maximiza a entropia da fuzzy-KAM. A Figura 14 mostra o gráfico da entropia da fuzzy-KAM em função do parâmetro α , obtida considerando o conjunto de memórias fundamentais \mathcal{A} definido no início deste exemplo e a medida de similaridade \mathcal{S}_H . Neste caso, temos que o valor de α que maximiza a entropia da fuzzy-KAM é $\alpha = 4,36$.

O terceiro passo é determinar a matriz $K = (\kappa_{ij})$, em que

$$\kappa_{ij} = e^{(\alpha * (\mathcal{S}_H(A^i, A^j) - 1))}.$$

Para este exemplo, temos:

$$K = \begin{pmatrix} 1 & 0,65 & 0,8 & 0,65 & 0,27 & 0,27 & 0,27 & 0,27 \\ 0,65 & 1 & 0,8 & 0,65 & 0,42 & 0,27 & 0,42 & 0,27 \\ 0,8 & 0,8 & 1 & 0,8 & 0,34 & 0,34 & 0,34 & 0,34 \\ 0,65 & 0,65 & 0,8 & 1 & 0,27 & 0,42 & 0,27 & 0,42 \\ 0,27 & 0,42 & 0,34 & 0,27 & 1 & 0,27 & 0,42 & 0,27 \\ 0,27 & 0,27 & 0,34 & 0,42 & 0,27 & 1 & 0,27 & 0,42 \\ 0,27 & 0,42 & 0,34 & 0,27 & 0,42 & 0,27 & 1 & 0,65 \\ 0,27 & 0,27 & 0,34 & 0,42 & 0,27 & 0,42 & 0,65 & 1 \end{pmatrix} \quad (5.33)$$

Em seguida, determinamos a matriz $G = K^{-1}$:

$$G = \begin{pmatrix} 2,83 & 0 & -2,28 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3,31 & -2,28 & -0,25 & -0,4 & 0,06 & -0,75 & 0,55 \\ -2,28 & -2,28 & 6,49 & -2,28 & 0 & 0 & 0 & 0 \\ 0 & -0,25 & -2,28 & 3,31 & 0,06 & -0,4 & 0,55 & -0,75 \\ 0 & -0,4 & 0 & 0,06 & 1,36 & -0,2 & -0,4 & 0,06 \\ 0 & 0,06 & 0 & -0,4 & -0,2 & 1,36 & 0,06 & -0,4 \\ 0 & -0,75 & 0 & 0,55 & -0,4 & 0,06 & 2,20 & -1,36 \\ 0 & 0,55 & 0 & -0,75 & 0,06 & -0,4 & -1,36 & 2,20 \end{pmatrix} \quad (5.34)$$

Dessa forma, com o parâmetro real $\alpha > 0$ obtido e a matriz G podemos definir uma fuzzy-KAM autoassociativa. A seguir, descreveremos a etapa de recordação, isto é, explicitaremos como determinar a saída da fuzzy-KAM a partir de um conjunto fuzzy de entrada X .

Consideremos o conjunto fuzzy de entrada $X = [0.4, 0.5]^T$. Primeiramente, determinamos o vetor $E(X) = [\kappa(A^1, X), \dots, \kappa(A^8, X)]^T$. Temos:

$$E(X) = [0,52 \quad 0,52 \quad 0,65 \quad 0,52 \quad 0,52 \quad 0,52 \quad 0,52 \quad 0,52]^T \quad (5.35)$$

Em seguida, calculamos os pesos

$$w_\xi = \frac{\sum_{\mu=1}^p g_{\xi\mu} \kappa(A^\mu, X)}{\sum_{\eta=1}^p \sum_{\mu=1}^p g_{\eta\mu} \kappa(A^\mu, X)}, \quad \xi = 1, \dots, 8.$$

Utilizando a forma matricial, temos que $\mathbf{w} = F(GE(X))$. Para este exemplo, temos:

$$\mathbf{w} = [0 \quad -0,15 \quad 0,58 \quad -0,15 \quad 0,22 \quad 0,22 \quad 0,14 \quad 0,14]^T \quad (5.36)$$

Finalmente, determinamos a saída Y da fuzzy-KAM calculando a combinação linear das memórias fundamentais A^ξ , $\xi = 1, \dots, 8$ com os pesos calculados anteriormente e aplicando em seguida a função φ . Obtemos:

$$Y = 0A^1 - 0,15A^2 + 0,58A^3 - 0,15A^4 + 0,22A^5 + 0,22A^6 + 0,14A^7 + 0,14A^8 = \begin{bmatrix} 0,4 \\ 0,5 \end{bmatrix}.$$

Exemplo 5. Consideremos agora o conjunto de memórias fundamentais

$\mathcal{A} = \{(A^1, B^1), (A^2, B^2), \dots, (A^8, B^8)\} \subset [0, 1]^2 \times [0, 1]^2$ formado pelos pares

$$\begin{aligned} (A^1, B^1) &= \left(\begin{bmatrix} 0, 1 \\ 0, 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), (A^2, B^2) = \left(\begin{bmatrix} 0, 2 \\ 0, 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), (A^3, B^3) = \left(\begin{bmatrix} 0, 2 \\ 0, 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), \\ (A^4, B^4) &= \left(\begin{bmatrix} 0, 2 \\ 0, 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right), (A^5, B^5) = \left(\begin{bmatrix} 0, 4 \\ 0, 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), (A^6, B^6) = \left(\begin{bmatrix} 0, 4 \\ 0, 8 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \\ (A^7, B^7) &= \left(\begin{bmatrix} 0, 6 \\ 0, 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), (A^8, B^8) = \left(\begin{bmatrix} 0, 6 \\ 0, 6 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \end{aligned}$$

a medida de similaridade \mathcal{S}_H . Definiremos uma fuzzy-KAM heteroassociativa e, em seguida, dado um conjunto fuzzy de entrada, descreveremos a etapa de recordação.

O primeiro passo é determinar a matriz $S = (s_{ij})$, em que $s_{ij} = \mathcal{S}_H(A^i, A^j)$. Em seguida, calculamos o valor do parâmetro α que maximiza a entropia da fuzzy-KAM. O terceiro passo é determinar a matriz $K = (\kappa_{ij})$, em que κ_{ij} é dado por (5.5) e, posteriormente, determinamos a matriz $G = K^{-1}$. Observemos que todos esses passos são realizados utilizando apenas a medida de similaridade e os conjuntos fuzzy A^ξ pertencentes às memórias fundamentais (A^ξ, B^ξ) . Uma vez que os conjuntos fuzzy A^ξ bem como a medida de similaridade que estamos considerando neste exemplo são iguais aos considerados no exemplo anterior, a matriz S , a matriz K e a matriz G são dadas por (5.32), (5.33) e (5.34) respectivamente e $\alpha = 4,36$.

Consideremos, como no exemplo anterior, o conjunto fuzzy de entrada $X = [0.4, 0.5]^T$. Primeiramente, determinamos o vetor $E(X) = [\kappa(A^1, X), \dots, \kappa(A^8, X)]^T$ e em seguida calculamos os pesos utilizando a forma matricial $\mathbf{w} = F(GE(X))$. Novamente, estas etapas, além do conjunto de entrada X , dependem apenas da medida de similaridade e dos conjuntos fuzzy A^ξ para serem calculadas. Logo, $E(X)$ e \mathbf{w} são dados por (5.35) e (5.36), respectivamente.

Finalmente, determinamos a saída Y da fuzzy-KAM heteroassociativa calculando a combinação linear dos conjuntos fuzzy B^ξ pertencentes às memórias fundamentais (A^ξ, B^ξ) , $\xi = 1, \dots, 8$, com os pesos calculados anteriormente e aplicando em seguida a função φ . Obtemos:

$$Y = 0B^1 - 0,15B^2 + 0,58B^3 - 0,15B^4 + 0,22B^5 + 0,22B^6 + 0,14B^7 + 0,14B^8 = \begin{bmatrix} 0,28 \\ 0,72 \end{bmatrix}.$$

Observemos que a matriz K definida por (5.5) tem dimensão $p \times p$, em que p denota o número de memórias fundamentais. Assim, o cálculo da matriz $G = K^{-1}$, conforme sugerido pelo Teorema 3, pode apresentar alto custo computacional, sobretudo nos casos em que o número de memórias fundamentais é grande. Uma possível estratégia

para amenizar este problema seria utilizar algum critério para a seleção de memórias fundamentais. Assim, a matriz K seria definida com base em um subconjunto do conjunto das memórias fundamentais e portanto teria menor dimensão. No entanto, existem outras estratégias que podem ser utilizadas. Por meio da formulação matricial da *fuzzy-KAM*, sugeriremos a seguir outra alternativa para diminuir o custo computacional exigido para o cálculo da matriz G .

Em vez de considerar na definição da *fuzzy-KAM* a matriz G sugerida pelo Teorema 3, a proposta é considerar uma matriz G com uma determinada estrutura esparsa. Ideias semelhantes vêm sendo exploradas recentemente em outros artigos da literatura como podemos ver em [32, 63, 77]. Especificamente, sugerimos considerar uma matriz G formada por blocos não nulos em torno da diagonal principal da seguinte forma:

$$G = \begin{pmatrix} G_1 & 0 & 0 & \cdots & 0 \\ 0 & G_2 & 0 & \cdots & 0 \\ 0 & 0 & G_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & G_M \end{pmatrix} \quad (5.37)$$

em que G_i , $i = 1, \dots, m$ são submatrizes não nulas de G e 0 denota uma submatriz nula. Seguindo a mesma ideia do Teorema 3, buscamos a matriz G com esta estrutura que melhor aproxima a inversa da matriz K dada por (5.5). Em outras palavras, denotando por $\|\cdot\|_F$ a norma de Frobenius de uma matriz e por I a matriz identidade, queremos determinar a matriz G que resolve o problema de quadrados mínimos

$$\min \|GK - I\|_F \quad (5.38)$$

e satisfaz a restrição na estrutura de G . Lembremos que a norma de Frobenius [23] de uma matriz $M = (m_{ij})$ de dimensão $m \times n$ é dada por:

$$\|M\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |m_{ij}|^2}.$$

Consideremos a matriz G com a estrutura em blocos dada por (5.37). Temos que:

$$GK = \begin{pmatrix} G_1 & 0 & 0 & \cdots & 0 \\ 0 & G_2 & 0 & \cdots & 0 \\ 0 & 0 & G_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & G_m \end{pmatrix} \begin{pmatrix} K_{11} & K_{12} & K_{13} & \cdots & K_{1m} \\ K_{21} & K_{22} & K_{23} & \cdots & K_{2m} \\ K_{31} & K_{32} & K_{33} & \cdots & K_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_{m1} & K_{m2} & K_{m3} & \cdots & K_{mm} \end{pmatrix} \quad (5.39)$$

$$= \begin{pmatrix} G_1 K_{11} & G_1 K_{12} & G_1 K_{13} & \cdots & G_1 K_{1m} \\ G_2 K_{21} & G_2 K_{22} & G_2 K_{23} & \cdots & G_2 K_{2m} \\ G_3 K_{31} & G_3 K_{32} & G_3 K_{33} & \cdots & G_3 K_{3m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_m K_{m1} & G_m K_{m2} & G_m K_{m3} & \cdots & G_m K_{mm} \end{pmatrix} \quad (5.40)$$

Logo:

$$\|GK - I\|_F^2 = \sum_{i=1}^m (\|G_i K_{ii} - I\|_F^2 + \sum_{\substack{j=1 \\ j \neq i}}^m \|G_i K_{ij}\|_F^2) = \sum_{i=1}^m \|G_i K_{i:} - J_i\|_F^2, \quad (5.41)$$

em que $K_{i:} = (K_{i1} K_{i2} \cdots K_{im})$ é a i -ésima linha de blocos da matriz K e $J_i = (0 \ 0 \cdots I \cdots 0)$ é uma matriz formada por uma linha de m blocos, sendo $m - 1$ nulos e o bloco da posição i igual a identidade. Assim, como todos os termos no último somatório de (5.41) são positivos, determinar G que minimiza $\|GK - I\|_F$ é equivalente a determinar m matrizes G_i que minimizam $\|G_i K_{i:} - J_i\|_F^2$. Logo, temos que:

$$G_i = J_i K_{i:}^\dagger, \quad i = 1, \dots, m, \quad (5.42)$$

onde $K_{i:}^\dagger$ denota a pseudo inversa da matriz $K_{i:}$ e m é o número de blocos.

Dessa forma, se necessário, podemos considerar a matriz G com a estrutura em blocos dada por (5.37), em que G_i é dada por (5.42).

Observação 11. *Observamos que essa forma alternativa de definir a matriz G requer um estudo mais aprofundado. Especificamente, precisa-se estabelecer um critério para definir o número de blocos e como tal divisão será feita. No entanto, em experimentos preliminares que serão apresentados no Capítulo 6, essa técnica se mostrou útil em problemas de reconhecimento de faces com um grande número de memórias fundamentais em que os blocos são definidos de maneira adequada usando informações adicionais sobre os indivíduos, como por exemplo gênero ou condições de iluminação.*

6 Aplicações em Problemas de Classificação

Em um problema de classificação o objetivo é associar um dado padrão X a um rótulo l que representa sua classe, com base em algumas de suas características. Geralmente, para realizar esta tarefa, utiliza-se um conjunto de padrões já rotulados, denominado conjunto de treinamento (ou conjunto das memórias fundamentais) dado por $A_{\mathcal{L}} = \{(A^\xi, l_\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{L}$, onde A^ξ , $\xi = 1, \dots, p$ são padrões distintos e l_ξ o rótulo que representa a classe de A^ξ . Neste capítulo, apresentaremos duas abordagens que permitem aplicar as *fuzzy*-KAMs em problemas de classificação de padrões e reconhecimento de faces. Especificamente, definiremos dois classificadores. O primeiro, inspirado pelos classificadores baseados em representação esparsa, utiliza uma abordagem autoassociativa enquanto que o segundo contempla o caso heteroassociativo.

Como frutos deste trabalho, ao longo do desenvolvimento da tese, foram produzidos um capítulo de livro [76], um artigo de revista [68] e quatro artigos apresentados em congressos [65–67, 69] que contém essas abordagens e alguns dos testes que descreveremos na Seção 6.4, bem como variações destes.

6.1 Abordagem Baseada em Memória Autoassociativa

Nesta seção definiremos um classificador com base na versão autoassociativa da *fuzzy*-KAM. Para facilitar a compreensão, explicaremos a seguir como tal classificador será definido. De forma geral, dado um problema de classificação composto pelo conjunto de treinamento $A_{\mathcal{L}} = \{(A^\xi, l_\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{L}$ e um padrão X a ser classificado, consideramos uma *fuzzy*-KAM autoassociativa definida com base no conjunto das memórias fundamentais $\{A^\xi, \xi = 1, \dots, p\} \subset \mathcal{F}(U)$. Em seguida, apresentamos X como padrão de entrada para a *fuzzy*-KAM e, a partir do padrão Y produzido como saída, dos pesos w_ξ calculados para definir a saída Y da *fuzzy*-KAM e do conjunto de treinamento, associamos X a um rótulo l , o qual indicará a classe a qual X pertence. A Figura 15 ilustra a abordagem baseada em memória autoassociativa.

A seguir, faremos uma breve revisão sobre classificadores baseados em representação esparsa [80], uma vez que a definição do classificador baseado na abordagem autoassociativa é inspirada por tais modelos.

Consideremos um universo de discurso finito U e um conjunto de rótulos $\mathcal{L} = \{l_1, \dots, l_m\}$. Seja $A_{\mathcal{L}} = \{(A^\xi, l_\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{L}$ um conjunto de padrões de treinamento, em que A^ξ são subconjuntos *fuzzy* distintos de U e l_ξ o rótulo que indica a que classe o padrão A^ξ pertence.

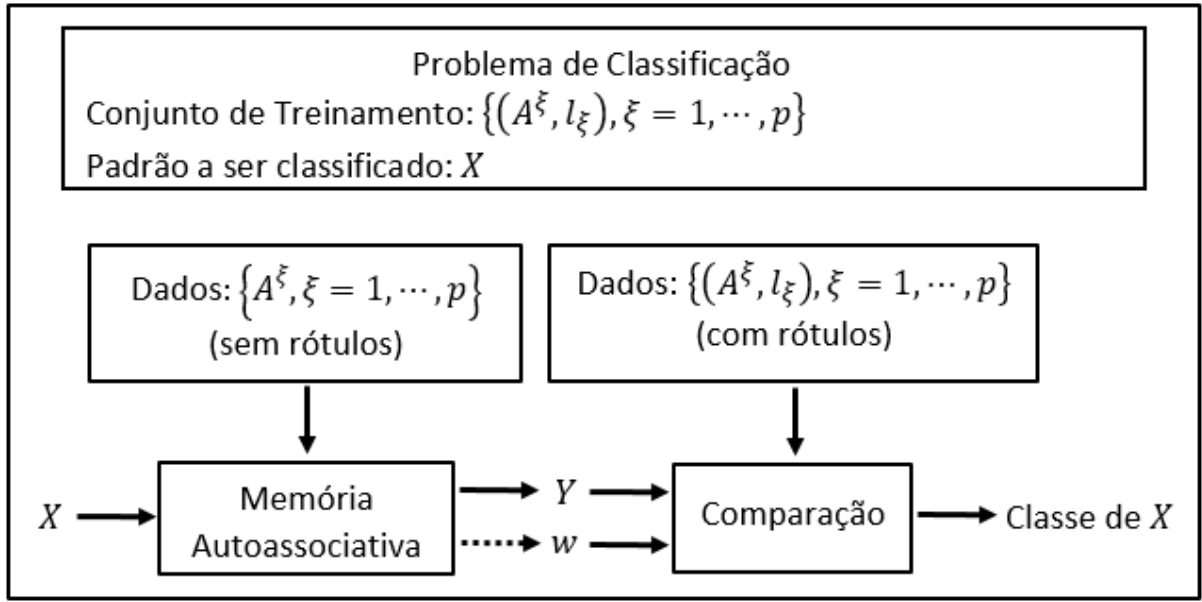


Figura 15 – Diagrama ilustrativo da abordagem baseada em memória autoassociativa.

Um classificador de representação esparsa [80] é definido com base na hipótese de que toda amostra Y de uma determinada classe i pode ser escrita aproximadamente como combinação linear dos padrões de treinamento desta classe i . Especificamente, se Y pertence a classe i , então:

$$Y(u) \approx \sum_{\xi \in \mathcal{L}_i} \alpha_\xi A^\xi(u), \quad \forall u \in U, \quad (6.1)$$

em que \mathcal{L}_i denota o conjunto dos índices $\xi \in \{1, \dots, p\}$ tais que $l_\xi = i$. Em termos matemáticos, $\mathcal{L}_i = \{\xi; l_\xi = i\} \subset \{1, \dots, p\}$. De modo equivalente, se definirmos $\alpha_\xi = 0$, para todo ξ tal que $l_\xi \neq i$, a amostra Y pode ser expressa como combinação linear de todos padrões de treinamento A^ξ , $\xi = 1, \dots, p$. Assim, temos que

$$Y(u) \approx \sum_{\xi=1}^p \alpha_\xi A^\xi(u), \quad \forall u \in U, \quad (6.2)$$

com $\alpha_\xi = 0$, para todo ξ tal que $l_\xi \neq i$.

Uma vez que os coeficientes α_ξ da combinação linear (6.2) são todos nulos, exceto por aqueles que correspondem a padrões da classe i , isto é, $\alpha_\xi = 0$ para todo ξ tal que $l_\xi \neq i$, temos que Y pode ser representado por meio de uma combinação linear esparsa dos padrões de treinamento A^ξ , $\xi = 1, \dots, p$.

Agora, consideremos uma *fuzzy*-KAM autoassociativa $\mathcal{K} : \mathcal{F}(U) \rightarrow \mathcal{F}(U)$ projetada para armazenar o conjunto de memórias fundamentais $\{A^1, \dots, A^p\}$. Consideremos também um padrão de entrada X e suponhamos que X pertença a uma determinada classe i (ou que X seja uma versão ruidosa de um padrão da classe i). Por um lado, espera-se

que a *fuzzy*-KAM, por se tratar de uma memória associativa, produza a partir de X um padrão $\mathcal{K}(X) = Y$ pertencente à mesma classe i . Por definição, $Y = \mathcal{K}(X)$ é dado por:

$$Y(u) = \varphi \left(\sum_{\xi=1}^p w_{\xi} A^{\xi}(u) \right), \quad (6.3)$$

onde $\varphi(x) = \max(0, \min(1, x))$. Por outro lado, inspirados pelos classificadores de representação esparsa, consideramos a hipótese de que Y pode ser escrito como combinação linear dos padrões de treinamento da classe i ou, de modo equivalente, como uma combinação linear esparsa dos padrões de treinamento. Assim, $\mathcal{K}(X) = Y$ também pode ser representado pela equação (6.2).

Comparando (6.2) e (6.3), exceto pela função φ , que pode ser desconsiderada se $\sum_{\xi=1}^p w_{\xi} A^{\xi}(u) \in [0, 1]$, concluímos que a combinação linear na equação (6.3) também deve ser esparsa. Portanto, os coeficientes α_{ξ} em (6.2) podem ser aproximados por

$$\alpha_{\xi} = w_{\xi} \chi_i(\ell_{\xi}), \quad \forall \xi = 1, \dots, p, \quad (6.4)$$

onde $\chi_i : \mathcal{L} \rightarrow \{0, 1\}$, para $i \in \mathcal{L}$, denota a função indicadora:

$$\chi_i(x) = \begin{cases} 1, & x = i, \\ 0, & \text{caso contrário.} \end{cases} \quad (6.5)$$

Observemos que (6.4) implica $\alpha_{\xi} = w_{\xi}$ se $\ell_{\xi} = i$ e $\alpha_{\xi} = 0$ caso contrário. Dessa forma, se o padrão de entrada X pertence a classe i , pressupomos que

$$Y(u) \approx \sum_{\xi=1}^p w_{\xi} \chi_i(\ell_{\xi}) A^{\xi}(u), \quad \forall u \in U. \quad (6.6)$$

No entanto, na prática, queremos atribuir ao padrão de entrada X um rótulo que representa sua classe. Portanto, a priori, não sabemos a que classe X pertence. Assim, a proposta é atribuir a X uma classe l que minimiza a distância entre Y e a combinação linear $\sum_{\xi=1}^p w_{\xi} \chi_l(\ell_{\xi}) A^{\xi}$. Precisamente, denotando por d_2 a distância L_2 , atribuímos a X uma classe $l \in \mathcal{L}$ tal que

$$d_2 \left(Y, \sum_{\xi=1}^p w_{\xi} \chi_l(\ell_{\xi}) A^{\xi} \right) \leq d_2 \left(Y, \sum_{\xi=1}^p w_{\xi} \chi_i(\ell_{\xi}) A^{\xi} \right), \quad \forall i \in \mathcal{L}. \quad (6.7)$$

O Algoritmo 1 resume a abordagem autoassociativa do classificador baseado na *fuzzy*-KAM.

Retomemos agora o Exemplo 4 para ilustrar a aplicação da abordagem baseada na versão autoassociativa da *fuzzy*-KAM em um problema de classificação.

Algoritmo 1 – Classificador baseado na fuzzy-KAM - Abordagem autoassociativa

Dados: Conjunto de treinamento $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \dots, p\} \subseteq \mathcal{F}(U) \times \mathcal{L}$,
medida de similaridade \mathcal{S} e parâmetro $\alpha > 0$;

Opcional: matriz $G \in \mathbb{R}^{p \times p}$ definida de acordo com o Teorema 3.

Entrada: $X \in \mathcal{F}(U)$.

Saída: Classe $\ell \in \mathcal{L}$.

se a matriz G não foi fornecida **então**

 | calcule-a a partir dos vetores A^1, \dots, A^p de acordo com o Teorema 3;

fim

Determine os pesos w_ξ , para $\xi = 1, \dots, p$, usando a equação (5.4);

Defina o conjunto fuzzy $Y \in \mathcal{F}(U)$ por meio da equação (5.3);

Escolha $\ell \in \mathcal{L}$ e calcule $\eta_\ell = d_2\left(Y, \sum_{\xi=1}^p w_{\xi 0} \chi_\ell(\ell_\xi) A^\xi\right)$;

para todo $i \in \mathcal{L} \setminus \{\ell\}$ **faça**

 | Calcule

$$\eta_i = d_2\left(Y, \sum_{\xi=1}^p w_{\xi 0} \chi_i(\ell_\xi) A^\xi\right); \quad (6.8)$$

se $\eta_i < \eta_\ell$ **então**

 | Atualize $\ell = i$ e $\eta_\ell = \eta_i$;

fim

fim

Exemplo 6. Consideremos o problema de classificação com duas classes cujo conjunto de treinamento é dado por

$$\{(A^1, 1), (A^2, 1), (A^3, 1), (A^4, 1), (A^5, 2), (A^6, 2), (A^7, 2), (A^8, 2)\} \subset [0, 1]^2 \times \{1, 2\},$$

em que os conjuntos fuzzy A^ξ , como no Exemplo 4, são dados por:

$$\begin{aligned} A^1 &= \begin{bmatrix} 0, 1 \\ 0, 5 \end{bmatrix}, & A^2 &= \begin{bmatrix} 0, 2 \\ 0, 4 \end{bmatrix}, & A^3 &= \begin{bmatrix} 0, 2 \\ 0, 5 \end{bmatrix}, & A^4 &= \begin{bmatrix} 0, 2 \\ 0, 6 \end{bmatrix}, \\ A^5 &= \begin{bmatrix} 0, 4 \\ 0, 2 \end{bmatrix}, & A^6 &= \begin{bmatrix} 0, 4 \\ 0, 8 \end{bmatrix}, & A^7 &= \begin{bmatrix} 0, 6 \\ 0, 4 \end{bmatrix}, & A^8 &= \begin{bmatrix} 0, 6 \\ 0, 6 \end{bmatrix} \end{aligned}$$

e o conjunto fuzzy a ser classificado é $X = [0, 4, 0, 5]^T$.

A Figura 16, extraída de [76], apresenta uma representação dos conjuntos fuzzy A^ξ pertencentes ao conjunto de treinamento bem como do conjunto fuzzy X a ser classificado. Os conjuntos pertencentes a classe 1 estão representados por quadrados vermelhos enquanto que os conjuntos pertencentes a classe 2 estão representados por triângulos azuis. O conjunto X está representado por um ponto preto.

Na abordagem baseada em memória autoassociativa, consideramos a fuzzy-KAM autoassociativa definida com base no conjunto de memórias fundamentais $\mathcal{A} =$

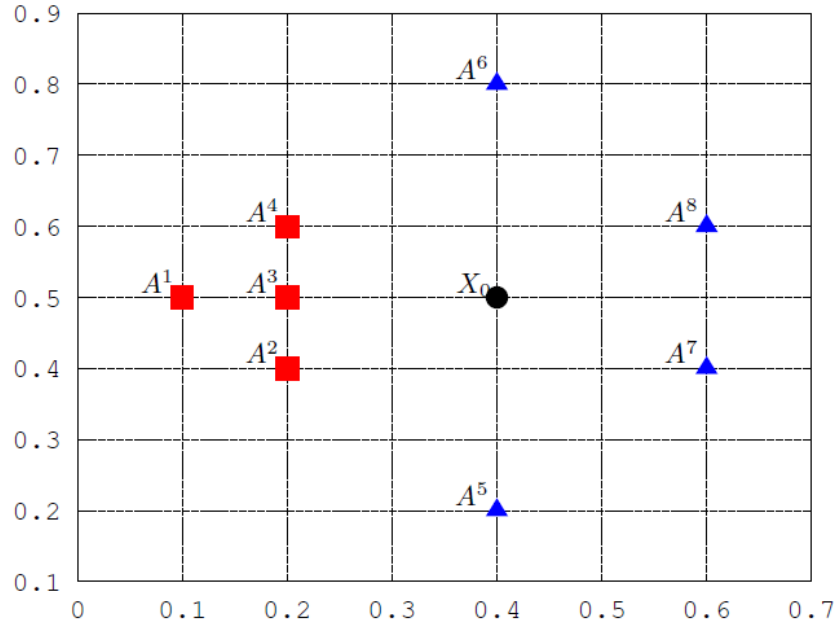


Figura 16 – Representação do conjunto de treinamento do problema de classificação considerado no Exemplo 6. Os conjuntos pertencentes a classe 1 estão representados por quadrados vermelhos enquanto que os conjuntos pertencentes a classe 2 estão representados por triângulos azuis. O conjunto X a ser classificado está representado por um ponto preto. Figura extraída de [76].

$\{A^1, \dots, A^8\} \subset [0, 1]^2$ obtido a partir do conjunto de treinamento e de uma medida de similaridade. Neste exemplo, como no Exemplo 4, consideraremos a medida de similaridade \mathcal{S}_H . O primeiro passo é apresentar o padrão X a ser classificado como entrada para esta fuzzy-KAM e determinar o conjunto fuzzy de saída Y . Conforme calculamos no Exemplo 4, ao apresentar $X = [0, 4, 0, 5]^T$ como padrão de entrada, a fuzzy-KAM autoassociativa produz $Y = [0, 4, 0, 5]^T$ como saída. Posteriormente, calculamos as combinações lineares $Y^l = \sum_{\xi=1}^p w_{\xi} \chi_l(l_{\xi}) A^{\xi}$, correspondentes a cada uma das classes $l = 1$ e $l = 2$. Lembremos que os pesos w_{ξ} também calculados no Exemplo 4 são dados por $\mathbf{w} = [0 \quad -0,15 \quad 0,58 \quad -0,15 \quad 0,22 \quad 0,22 \quad 0,14 \quad 0,14]^T$. Assim, obtemos:

$$Y^1 = 0A^1 - 0,15A^2 + 0,58A^3 - 0,15A^4 + 0A^5 + 0A^6 + 0A^7 + 0A^8 = [0,06, 0,14]^T$$

e

$$Y^2 = 0A^1 + 0A^2 + 0A^3 + 0A^4 + 0,22A^5 + 0,22A^6 + 0,14A^7 + 0,14A^8 = [0,34, 0,36]^T.$$

Finalmente, calculamos as distâncias entre Y e as combinações lineares Y^1 e Y^2 . Temos que $d_2(Y, Y^1) = 0,49$ e $d_2(Y, Y^2) = 0,15$. Como $d_2(Y, Y^2) < d_2(Y, Y^1)$, associamos X à classe 2.

6.2 Abordagem Baseada em Memória Heteroassociativa

Nesta seção, definiremos um classificador com base na versão heteroassociativa da *fuzzy*-KAM.

Consideremos um problema de classificação baseado no conjunto de treinamento $A_{\mathcal{L}} = \{(A^\xi, l_\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \mathcal{L}$ e a *fuzzy*-KAM heteroassociativa projetada para armazenar o seguinte conjunto de memórias fundamentais $\{(A^\xi, B^\xi), \xi = 1, \dots, p\} \subset \mathcal{F}(U) \times \{0, 1\}^m$, onde $m = \text{Card}(\mathcal{L})$ denota o número de classes, A^ξ representa uma amostra de uma determinada classe e $B^\xi \subset \{0, 1\}^m$, obtido com base em l_ξ , indica a classe a qual A^ξ pertence. Em termos matemáticos, se o padrão A^ξ pertence a classe i , isto é, se $l_\xi = i$, o conjunto (*fuzzy*) B^ξ associado à A^ξ é definido por:

$$B^\xi(v) = \begin{cases} 1, & \text{se } v = i \\ 0, & \text{se } v \neq i. \end{cases} \quad (6.9)$$

Dado um padrão X a ser classificado, o primeiro passo é determinar o conjunto *fuzzy* produzido pela *fuzzy*-KAM, isto é, calculamos $Y = \mathcal{K}(X)$ por meio da equação:

$$Y(v) = \varphi \left(\sum_{\xi=1}^p w_\xi^A B^\xi(v) \right), \quad \forall v \in V.$$

Observa-se que, devido a forma como os vetores B^ξ foram definidos, a menos da função φ que pode ser desconsiderada se $\sum_{\xi=1}^p w_\xi B^\xi(u) \in [0, 1]$, a componente $Y(i)$ corresponde à soma dos pesos w_ξ associados aos padrões pertencentes à classe i . Assim, temos que $Y(i)$ pode ser interpretado como a possibilidade de que o padrão de entrada pertença a classe i . Por este motivo, associamos o padrão de entrada X à i -ésima classe, em que i é o primeiro índice tal que $Y(i) \geq Y(j)$, para todo $j = 1, \dots, p$. A Figura 17 ilustra a abordagem baseada em memória heteroassociativa.

O Algoritmo 2 resume a abordagem heteroassociativa do classificador baseado na *fuzzy*-KAM.

Retomemos agora o Exemplo 5 para ilustrar a aplicação da abordagem baseada na versão heteroassociativa da *fuzzy*-KAM em um problema de classificação.

Exemplo 7. Consideremos o mesmo problema de classificação do Exemplo 6. Recordemos que conjunto de treinamento é dado por

$$\{(A^1, 1), (A^2, 1), (A^3, 1), (A^4, 1), (A^5, 2), (A^6, 2), (A^7, 2), (A^8, 2)\} \subset [0, 1]^2 \times \{1, 2\}$$

em que os conjuntos *fuzzy* A^ξ são dados por:

$$A^1 = \begin{bmatrix} 0, 1 \\ 0, 5 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0, 2 \\ 0, 4 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 0, 2 \\ 0, 5 \end{bmatrix}, \quad A^4 = \begin{bmatrix} 0, 2 \\ 0, 6 \end{bmatrix},$$

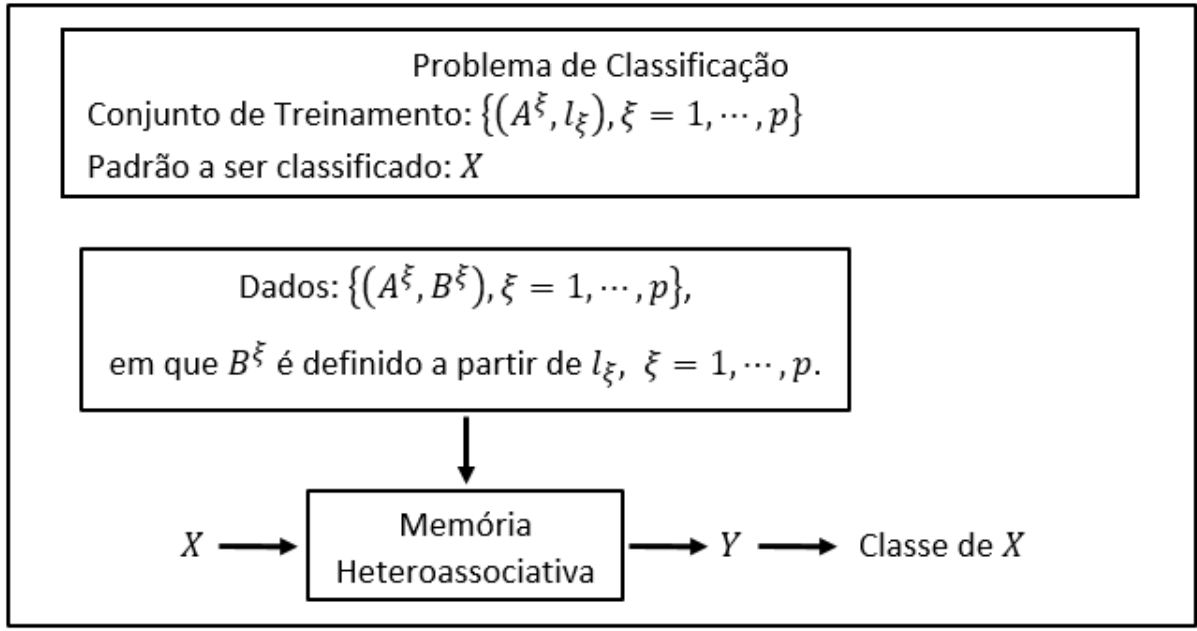


Figura 17 – Diagrama ilustrativo da abordagem baseada em memória heteroassociativa.

Algoritmo 2 – Classificador baseado na *fuzzy*-KAM - Abordagem heteroassociativa

Dados: Conjunto de treinamento $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \dots, p\} \subseteq \mathcal{F}(U) \times \mathcal{L}$,
medida de similaridade \mathcal{S} e parâmetro $\alpha > 0$;

Opcional: matriz $G \in \mathbb{R}^{p \times p}$ definida de acordo com o Teorema 3.

Entrada: $X \in \mathcal{F}(U)$.

Saída: Classe $\ell \in \mathcal{L}$.

se a matriz G não foi fornecida **então**

 | calcule-a a partir dos vetores A^1, \dots, A^p de acordo com o Teorema 3;

fim

Determine os pesos w_ξ , para $\xi = 1, \dots, p$, usando a equação (5.4);

Determine o conjunto *fuzzy* $Y \in \mathcal{F}(U)$ por meio da equação (5.3);

Defina $\ell = 1$

para *todo* $i = 1, \dots, m - 1$ **faça**

se $Y(i + 1) > Y(i)$ **então**

 | Atualize $\ell = i + 1$

fim

fim

$$A^5 = \begin{bmatrix} 0, 4 \\ 0, 2 \end{bmatrix}, \quad A^6 = \begin{bmatrix} 0, 4 \\ 0, 8 \end{bmatrix}, \quad A^7 = \begin{bmatrix} 0, 6 \\ 0, 4 \end{bmatrix}, \quad A^8 = \begin{bmatrix} 0, 6 \\ 0, 6 \end{bmatrix}$$

e, o padrão a ser classificado é $X = [0, 4, 0, 5]^T$. Neste problema de classificação temos duas classes, denotadas por 1 e 2. A partir desses rótulos, definimos os conjuntos B^ξ conforme (6.9). Dessa forma, como os padrões A^ξ , $\xi = 1, \dots, 4$, pertencem à classe 1 e os

padrões A^ξ , $\xi = 5, \dots, 8$, pertencem à classe 2, temos que:

$$B^\xi = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \xi = 1, \dots, 4 \quad e \quad B^\xi = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \xi = 5, \dots, 8.$$

Dessa forma, a partir do conjunto de treinamento dado, obtemos um conjunto de memórias fundamentais $\{(A^\xi, B^\xi), \xi = 1, \dots, 8\} \subset [0, 1]^2 \times \{0, 1\}^2$. Observemos que este conjunto de memórias fundamentais coincide com o considerado no Exemplo 5.

Na abordagem baseada em memória heteroassociativa, consideramos a fuzzy-KAM heteroassociativa definida com base no conjunto de memórias fundamentais $\{(A^\xi, B^\xi), \xi = 1, \dots, 8\} \subset [0, 1]^2 \times \{0, 1\}^2$ obtido a partir do conjunto de treinamento e de uma medida de similaridade. Neste exemplo, também consideraremos a medida de similaridade \mathcal{S}_H .

O primeiro passo é apresentar o padrão $X = [0, 4, 0, 5]^T$ a ser classificado como entrada para esta fuzzy-KAM e determinar o conjunto fuzzy de saída Y . Ao apresentar X como padrão de entrada, conforme o Exemplo 5, a fuzzy-KAM heteroassociativa produz $Y = [0, 28, 0, 72]^T$ como saída. Assim, como $Y(2) = 0, 72 > 0, 28 = Y(1)$, associamos X à classe 2.

6.3 Sobre o parâmetro α

Com o objetivo de tornar o modelo mais flexível, investigamos o efeito de considerar, não apenas um parâmetro α , mas vários parâmetros na definição dos classificadores. Assim, além da proposta original com apenas um parâmetro, consideramos outros três casos.

A primeira alternativa foi considerar um parâmetro α diferente para cada classe. Posteriormente, consideramos um parâmetro diferente para cada memória fundamental e finalmente, avaliamos o efeito de considerar um parâmetro para cada componente de cada memória fundamental. No entanto, nos testes realizados, não houve melhora nos resultados com o aumento no número de parâmetros considerados. Além disso, o ajuste de um maior número de parâmetros provocou um grande aumento do custo computacional, tornando essas alternativas inviáveis, sobretudo nos problemas com grande número de memórias fundamentais ou com memórias fundamentais de grande dimensão. Por este motivo, optamos por considerar apenas um parâmetro, como proposto inicialmente.

6.4 Experimentos Computacionais

Nesta seção apresentaremos os diversos experimentos computacionais realizados para avaliar o desempenho de ambos os classificadores baseados na fuzzy-KAM. Utilizamos

medidas de similaridade baseadas em cardinalidade, a medida de similaridade estrutural e medidas que satisfazem a versão normalizada da definição proposta por Xuecheng dada por (12), as quais foram apresentadas no Capítulo 2. Consideramos diversos problemas de classificação disponíveis nos repositórios *Knowledge Extraction Based on Evolutionary Learning* (KEEL) [4] etambém na base de dados da UCI [6]. Para os problemas de reconhecimento de faces consideramos as bases de imagens faciais AR [49], ORL [60] e GT [1]. Os experimentos realizados serão descritos detalhadamente a seguir.

6.4.1 Alguns Problemas de Classificação das Bases KEEL e UCI

Realizamos experimentos computacionais com vinte e dois problemas disponíveis nos repositórios *Knowledge Extraction Based on Evolutionary Learning* (KEEL) [4] e da UCI (*University of California, Irvine, Dept. of Information and Computer Sciences repository of machine learning databases*) [6], a saber: appendicitis, cleveland [15], crx, ecoli, glass, heart, iris, monks, movementlibras, pima, sonar, spectfheart, vowel, wdbc, wine, satimage, texture, german, yeast, spambase, phoneme e page-blocks [4, 6]. Este experimento foi apresentado em [67].

Estes experimentos foram realizados usando validação cruzada com 10 pastas conforme experimentos anteriores descritos na literatura [3]. Este método consiste em particionar o conjunto de dados em 10 partes e realizar 10 testes, utilizando em cada um deles uma das partes como conjunto de teste e as outras nove restantes como conjunto de treinamento. Posteriormente, calculamos a acurácia média de classificação com base nos valores de acurácia obtidos em cada um dos dez testes. A acurácia de classificação, ou porcentagem de acertos, é dada pela razão entre o número de padrões de testes classificados corretamente pelo classificador e o número total de padrões de testes considerado. Observamos que foi utilizado o mesmo particionamento dos dados que em [3, 4] para garantir uma comparação justa.

Alguns dos conjuntos considerados nestes experimentos contém, além dos atributos numéricos, atributos categóricos. Por este motivo, como descrito no experimento anterior, aqui também foi necessário realizar uma etapa de pré-processamento, na qual os dados originais foram convertidos em conjuntos *fuzzy*. O primeiro passo foi transformar cada atributo categórico $f \in \{v_1, \dots, v_c\}$, com $c > 1$, em um atributo numérico c -dimensional $\mathbf{n} = (n_1, n_2, \dots, n_c) \in \mathbb{R}^c$ da seguinte forma, para todo $i = 1, \dots, c$:

$$n_i = \begin{cases} 1, & \text{se } f = v_i, \\ 0, & \text{caso contrário,} \end{cases} \quad (6.10)$$

Como exemplo podemos citar o conjunto de dados crx. Um de seus atributos categóricos, o qual possui 14 possibilidades, foi transformado em 14 atributos numéricos por meio de (6.10).

Ao final do processo, em vez de 9 atributos categóricos e 6 numéricos do conjunto original, uma amostra do conjunto de dados crx transformado contém 46 atributos numéricos.

A conversão dos atributos categóricos em atributos numéricos possibilitou a representação de cada amostra de um conjunto de dados como um par (\mathbf{x}, ℓ) , onde $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ é um vetor numérico e $\ell \in \mathcal{L}$ denota sua classe. Em seguida, cada vetor de atributos $\mathbf{x} \in \mathbb{R}^n$ foi associado a um conjunto *fuzzy* $A = [a_1, a_2, \dots, a_n]^T$ por meio da equação:

$$a_i = \frac{1}{1 + e^{-(x_i - \mu_i)/\sigma_i}} \in [0, 1], \quad \forall i = 1, \dots, n, \quad (6.11)$$

onde μ_i e σ_i representam respectivamente a média e o desvio padrão da i -ésima componente de todos os padrões de treinamento. Dessa forma, cada amostra foi associada a um par (A, ℓ) , em que A é um conjunto *fuzzy* e ℓ denota sua classe. Consequentemente, o conjunto de treinamento pode ser escrito como uma família de conjuntos *fuzzy* rotulados $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \dots, p\}$.

Dado o conjunto de treinamento $\mathcal{A}_{\mathcal{L}} = \{(A^\xi, \ell_\xi) : \xi = 1, \dots, p\}$, devido a hipótese de que $A^\xi \neq A^\eta$, se $\xi \neq \eta$, no Teorema 3, foi necessário remover alguns elementos repetidos do conjunto de treinamento dos problemas texture, yeast, spambase, phoneme e page-blocks.

Consideramos os classificadores baseados nas abordagens autoassociativa e heteroassociativa da *fuzzy*-KAM com a medida de similaridade de Gregson, dada pela equação (2.9) e o parâmetro α que maximiza a entropia. A medida de similaridade de Gregson, além de satisfazer a definição proposta por Xuecheng, é uma medida de similaridade forte baseada em cardinalidade, obtida considerando a t-norma do mínimo nos esquemas de fuzzificação dados pelas equações (2.4), (2.5), apresentadas no Capítulo 2, e W -transitiva, em que W denota a t-norma de Łukasiewicz.

A Figura 18, extraída de [67], mostra os *boxplot* das acurácias de classificação médias obtidas pelos classificadores baseados nas abordagens autoassociativas e heteroassociativas da *fuzzy*-KAMs bem como as acurácias obtidas por outros nove modelos da literatura, a saber: 2SLAVE [24], FH-GBML [35], SGERD [48], CBA [45], CBA2 [46], CMAR [44], CPAR [82], C4.5 [59], e FARC-HD [3]. A acurácia dos nove primeiros classificadores foi extraída de [3]. Lembremos que a acurácia de classificação é dada pela razão entre o número de padrões de teste classificados corretamente e o número total de padrões de teste.

Podemos observar na Figura 18 que os classificadores baseados nas *fuzzy*-KAMs obtiveram desempenho satisfatório. De fato, tais classificadores superaram ou apresentaram resultado competitivo com os outros classificadores da literatura. Os dois *outliers* nos *boxplot* dos classificadores baseados nas *fuzzy*-KAMs correspondem aos problemas cleveland e yeast.

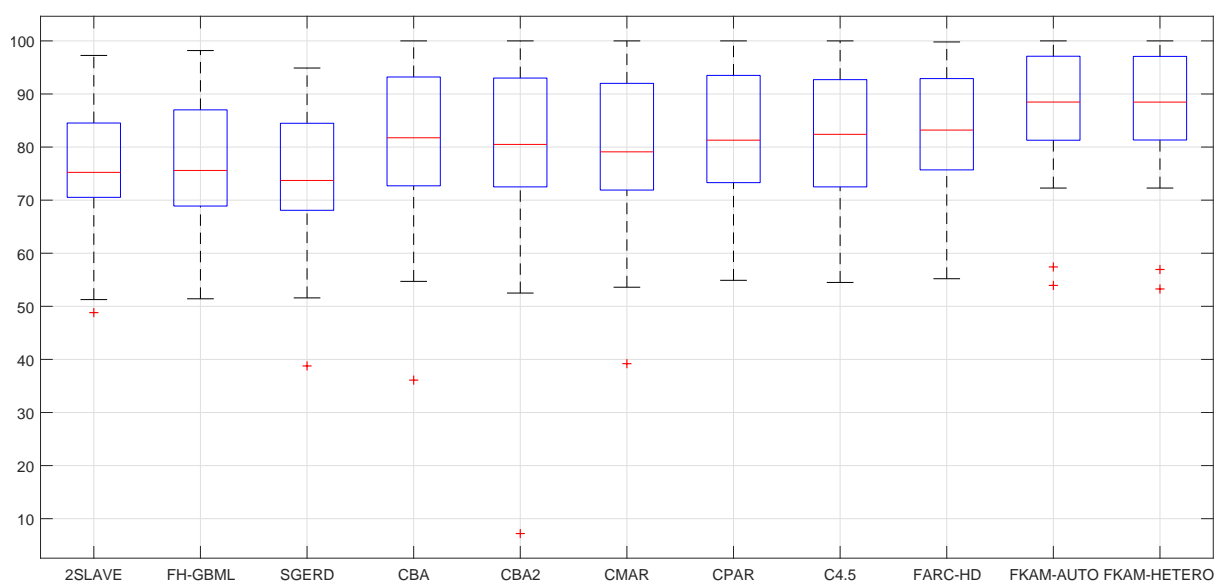


Figura 18 – *Boxplot* das acurácias de classificação de vários modelos da literatura em vinte e dois problemas. A acurácia dos nove primeiros classificadores foram extraídos de [3]. Figura extraída de [67].

A Tabela 3 mostra a acurácia de classificação obtida por cada modelo em cada um dos problemas.

Tabela 3 – Acurácias de classificação obtidas por diversos modelos em 22 problemas de classificação. A acurácia dos nove primeiros classificadores foram extraídos de [3].

	2SLAVE	FH-GBML	SGERD	CBA	CBA2	CMAR	CPAR	C4.5	FARC-HD	FKAM-AUTO	FKAM-HETERO
Appendicitis	82,91	86,00	84,48	89,60	89,60	89,70	87,80	83,30	84,20	86,00	86,00
Cleveland	48,82	53,51	51,59	56,90	54,90	53,90	54,90	54,50	55,20	53,94	53,27
Crx	74,06	86,60	85,03	83,60	85,00	85,00	87,30	85,30	86,00	84,81	84,81
Ecoli	84,53	69,38	74,05	78,00	77,10	77,70	76,20	79,50	82,20	83,68	83,39
Glass	58,05	57,99	58,49	70,80	71,30	70,30	68,90	67,40	70,20	81,29	81,33
Heart	71,36	75,93	73,21	83,00	81,50	82,20	80,70	78,50	84,40	82,96	82,59
Iris	94,44	94,00	94,89	93,30	93,30	94,00	96,00	96,00	96,00	93,33	93,33
Monks	97,26	98,18	80,65	100,00	100,00	100,00	100,00	100,00	99,80	100,00	100,00
MovementLibras	67,04	68,89	68,09	36,10	7,20	39,20	63,60	69,40	76,70	88,06	88,06
Pima	73,71	75,26	73,37	72,70	72,50	75,10	74,50	74,00	75,70	72,28	72,28
Sonar	71,42	68,24	71,90	75,40	77,90	78,80	75,00	70,50	80,20	88,90	88,88
Spectfheart	79,17	72,36	78,16	79,80	79,80	79,40	78,30	76,50	79,80	80,56	81,31
Vowel	71,11	67,07	65,83	63,60	74,90	60,40	63,00	81,50	71,80	99,39	99,39
Wdbc	92,33	92,26	90,68	94,70	95,10	94,90	95,10	95,20	95,30	97,37	97,72
Wine	89,47	92,61	91,88	93,80	93,80	96,70	95,60	93,30	94,30	97,75	98,30
SatImage	81,69	74,72	77,10	85,20	83,60	84,90	85,80	85,80	87,30	92,48	92,54
Texture	81,57	70,15	71,66	84,50	52,50	73,80	90,70	92,60	92,90	99,51	99,55
German	70,53	87,01	67,97	75,30	73,70	71,90	73,30	72,50	72,80	75,40	74,80
Yeast	51,27	51,42	38,77	54,70	55,90	53,60	56,30	55,60	58,50	57,42	56,94
Spambase	70,14	77,22	72,98	93,20	93,00	92,00	93,50	92,70	91,90	93,26	92,71
Phoneme	76,41	79,66	75,55	80,50	81,20	78,60	81,90	86,80	82,10	90,23	90,17
Page-Blocks	91,39	94,21	90,72	94,00	95,90	94,40	96,10	97,10	95,00	97,11	97,08

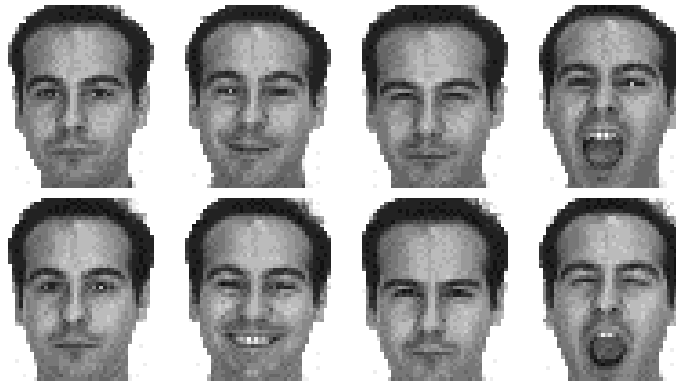


Figura 19 – Imagens de um indivíduo da base AR pertencente ao conjunto de treinamento.

6.4.2 Problemas de Reconhecimento de Faces

Um problema de reconhecimento de faces é um tipo particular de problema de classificação cujo objetivo é identificar a pessoa em uma imagem facial. Tal tarefa é realizada com base em um conjunto de imagens rotuladas, denominado conjunto de treinamento. Nos experimentos desenvolvidos, consideramos essencialmente três bases de imagens faciais, a saber: a base de imagens faciais AR [49], a base de imagens ORL [60] e a base GT [1].

1. Experimentos com a Base de Imagens Faciais AR

A base de imagens faciais AR [49] é composta por imagens em tons de cinza de 120 indivíduos com diversas expressões faciais, diferentes condições de iluminação, além de imagens dos indivíduos com óculos escuros e cachecol (os mesmos óculos escuros e cachecol para todas as pessoas). Tal como realizado por Luo et al. [47], as imagens da base AR foram cortadas e redimensionadas para dimensão 50×40 .

Com o objetivo de comparar os resultados com o estado da arte, realizamos os mesmos experimentos apresentados em Luo et al [47], conforme descrito a seguir. Consideramos 8 imagens de cada indivíduo do grupo de 120 pessoas como conjunto de treinamento. A Figura 19 mostra as imagens de treinamento consideradas para um indivíduo.

Considerando o conjunto de treinamento descrito acima, realizamos dois tipos de experimentos com a base de dados AR. No primeiro, ao qual nos referiremos como experimento a), consideramos o conjunto de teste composto por 4 imagens de cada indivíduo com óculos escuros e diferentes condições de iluminação. No segundo experimento, denominado experimento b), consideramos o conjunto de teste composto por 4 imagens de cada indivíduo com cachecol e diferentes condições de iluminação.

A Figura 20 mostra as imagens de teste do indivíduo da Figura 19 nos experimentos: a) óculos escuros + iluminação e b) cachecol + iluminação, respectivamente.



a) óculos escuros + iluminação.



b) cachecol + iluminação.

Figura 20 – Imagens do conjunto de teste do indivíduo da Figura 19 da base AR nos experimentos a) e b).

Inicialmente, consideramos a versão heteroassociativa do classificador baseado na *fuzzy*-KAM e as medidas de similaridade baseadas em cardinalidade. O objetivo deste experimento, apresentado em [68], foi avaliar a influência da escolha da medida de similaridade no desempenho da *fuzzy*-KAM. Em [68], o classificador foi definido utilizando a abordagem heteroassociativa baseada em um único passo de uma GEB-FAM. Lembremos que, como mencionado anteriormente, a *fuzzy*-KAM e a GEB-FAM de passo único são modelos equivalentes.

Recordando, a classe de medidas de similaridade baseadas em cardinalidade propostas por De Baets e De Meyer em [12] são definidas com base na expressão:

$$S(A, B) = \frac{a\alpha_{A,B} + b\omega_{A,B} + c\delta_{A,B} + d\nu_{A,B}}{a'\alpha_{A,B} + b'\omega_{A,B} + c'\delta_{A,B} + d'\nu_{A,B}}. \quad (6.12)$$

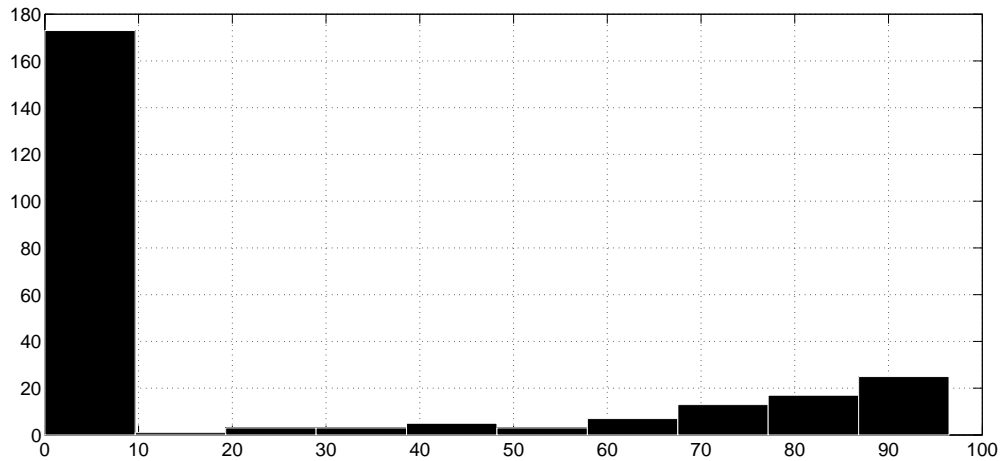
onde $a, a', b, b', c, c', d, d' \in \{0, 1\}$ são parâmetros e

- (i) $\alpha_{A,B} = \min\{\#(A \setminus B), \#(B \setminus A)\},$
- (ii) $\omega_{A,B} = \max\{\#(A \setminus B), \#(B \setminus A)\},$
- (iii) $\delta_{A,B} = \#(A \cap B),$
- (iv) $\nu_{A,B} = \#(A \cup B)^c.$

A Tabela 1, apresentada no Capítulo 2, mostra algumas medidas de similaridade obtidas a partir da equação (6.12).

Este experimento foi realizado considerando todas as medidas de similaridade listadas na Tabela 1. Para cada uma delas, utilizamos os esquemas de fuzzificação descritos

a) óculos escuros + iluminação



b) cachecol + iluminação

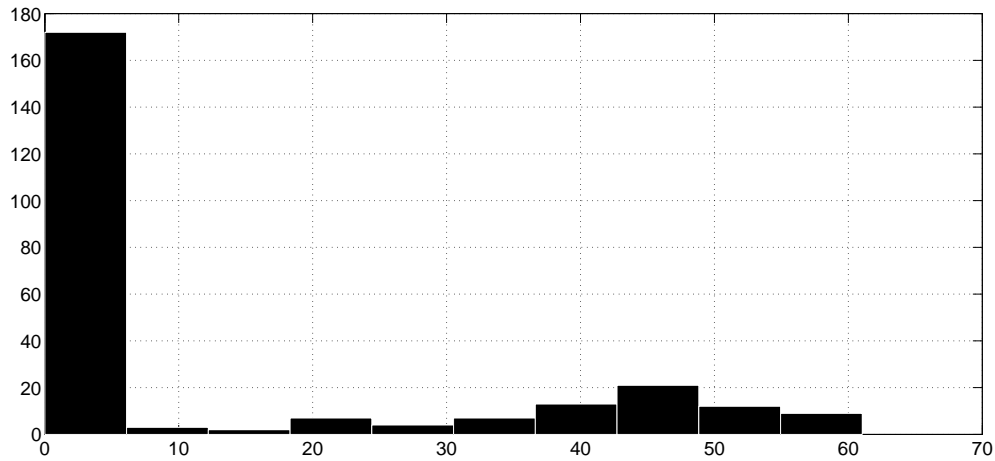


Figura 21 – Histograma das taxas de acerto obtidas considerando todas as 250 *fuzzy*-KAMs. Figuras extraídas de [68]

pelas equações (2.4), (2.5), (2.6) e (2.7) apresentadas no Capítulo 2. Lembremos que esses esquemas são todos baseados em t-normas de Frank. Especificamente, focamos nas t-normas do mínimo, do produto e na t-norma de Łukasiewicz. Por meio desse procedimento, obtivemos três versões de cada uma das medidas $R_1, R_2, \dots, R_{10}, R_{12}, \dots, R_{15}, S_{17}$ e S_{18} . Além dessas, consideramos também as medidas R_{11} e S_{19} , as quais dependem apenas de cardinalidade dos conjuntos *fuzzy* envolvidos. Dessa forma, obtivemos 50 medidas de similaridade. Para cada uma delas, consideramos cinco valores para o parâmetro α , a saber, $\alpha \in \{0.5, 1, 5, 10, 30\}$. Assim, combinando as 50 medidas de similaridade com os cinco valores do parâmetro α obtivemos 250 diferentes *fuzzy*-KAMs no total.

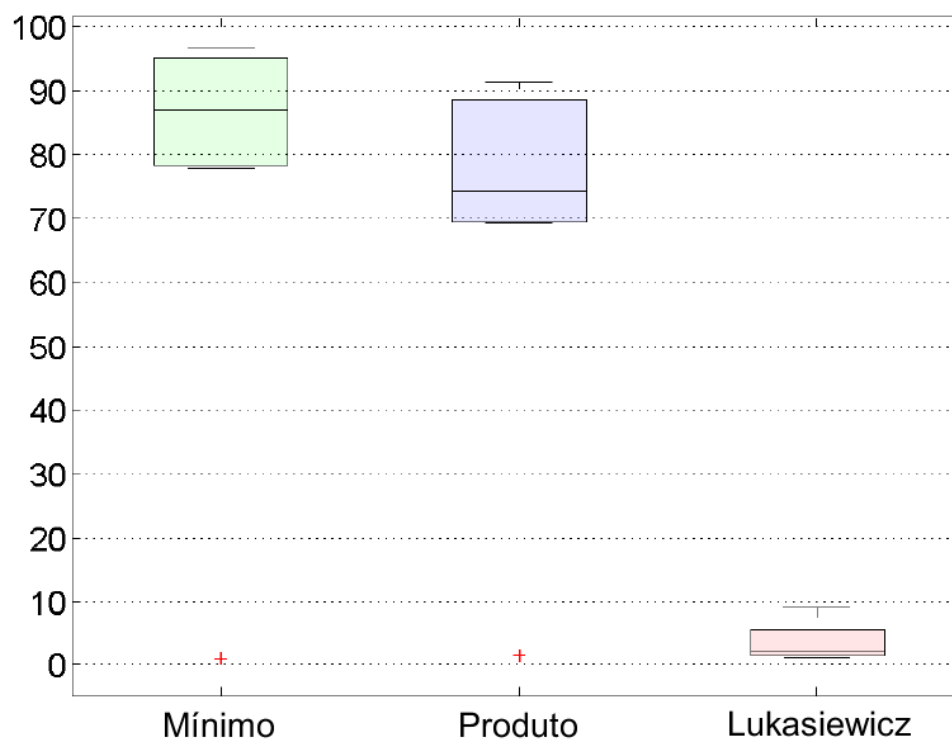
A Figura 21 apresenta um histograma das taxas de acerto obtidas pelas 250 *fuzzy*-KAMs nos experimentos a) e b).

Como podemos observar, grande parte das *fuzzy*-KAMs produziram taxas de acertos menores que ou iguais a 10%. A maioria dessas baixas taxas de acerto foi produzida por *fuzzy*-KAMs baseadas em medidas de similaridades não T -transitivas com respeito às t -normas do mínimo, produto, drástica e a t -norma de Łukasiewicz. De fato, dentre as *fuzzy*-KAMs definidas com medidas de similaridade não T -transitivas, apenas a baseada na medida S_{17} produziu taxa de reconhecimento maior que 10%. Precisamente, a *fuzzy*-KAM definida com a medida localmente reflexiva S_{17} baseada na t -norma do mínimo produziu uma acurácia de 94.58% (para $\alpha = 30$) e 53.54% (para $\alpha = 5$) nos experimentos a) e b), respectivamente. Por este motivo, nos concentramos nas 24 medidas de similaridade T -transitivas definidas a partir dos esquemas de fuzzificação dados pelas equações (2.4), (2.5), (2.6) e (2.7) baseados nas t -normas do mínimo, do produto e na t -norma de Łukasiewicz. Especificamente, nos concentramos nas três versões de cada uma das seguintes medidas R_i , para $i \in \{1, 2, 5, 6, 13, 14, 15\}$ e S_{18} .

Em seguida, para cada uma dessas medidas T -transitivas, selecionamos o parâmetro $\alpha \in \{0.5, 1, 5, 10, 30\}$ que produziu a maior taxa de reconhecimento. A Figura 22 mostra o *boxplot* dos valores da acurácia obtidos pelos 24 classificadores definidos utilizando estas medidas pelo esquema de fuzzificação. O primeiro *boxplot* foi construído com os resultados das *fuzzy*-KAMs com o parâmetro α que produziu o melhor resultado e as medidas de similaridades T -transitivas definidas usando a t -norma do mínimo nas equações (2.4), (2.5), (2.6) e (2.7). Analogamente, o segundo e o terceiro *boxplot* representam às melhores *fuzzy*-KAMs definidas com medidas de similaridade T -transitivas baseadas nas t -normas do produto e de Łukasiewicz respectivamente. Como podemos observar na Figura 22, as *fuzzy*-KAMs obtidas considerando a t -norma de Łukasiewicz também produziram baixas taxas de reconhecimento. No entanto, exceto pelas *fuzzy*-KAMs baseadas na medida de similaridade R_{13} , a qual correspondem os dois *outliers* presentes nos *boxplot* na Figura 22, as *fuzzy*-KAMs baseadas nas medidas de similaridade obtidas com as t -normas do mínimo e do produto obtiveram um desempenho satisfatório e competitivo com o obtido por outros modelos da literatura. Além disso, conforme mencionado na Observação 2, temos que, além de T -transitivas, as medidas R_i , para $i \in \{1, 2, 5, 6, 13, 14, 15\}$, são reflexivas e a medida S_{18} é localmente reflexiva quando consideramos $T = M$. O fato dos classificadores baseados na abordagem heteroassociativa terem obtido melhores resultados, em ambos os experimentos a) e b), quando consideramos medidas T -transitivas baseadas na t -norma do mínimo sugere que, além da T -transitividade, a reflexividade é uma propriedade importante a ser considerada.

Realizamos também os experimentos a) e b) com o classificador definido com base na abordagem autoassociativa, bem como considerando o classificador definido com base na abordagem heteroassociativa com outras medidas de similaridade. Agora, consideramos o parâmetro α que maximiza a entropia e as medidas de similaridade de Gregson, \mathcal{S}_G , a medida de similaridade de Eisler e Ekman, \mathcal{S}_E , a medida baseada na

a) óculos escuros+iluminação



b) cachecol+iluminação

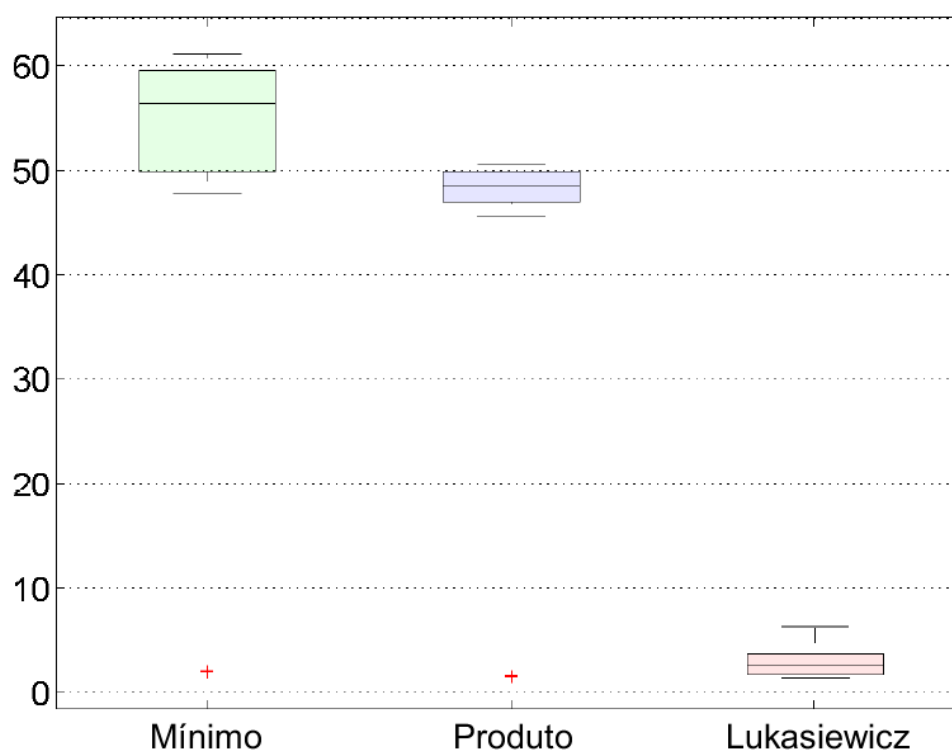


Figura 22 – *Boxplot* das acurácias produzidas pelas *fuzzy*-KAMs baseadas em medidas de similaridade T-transitivas pelo esquema de fuzzificação. Figuras extraídas de [68].

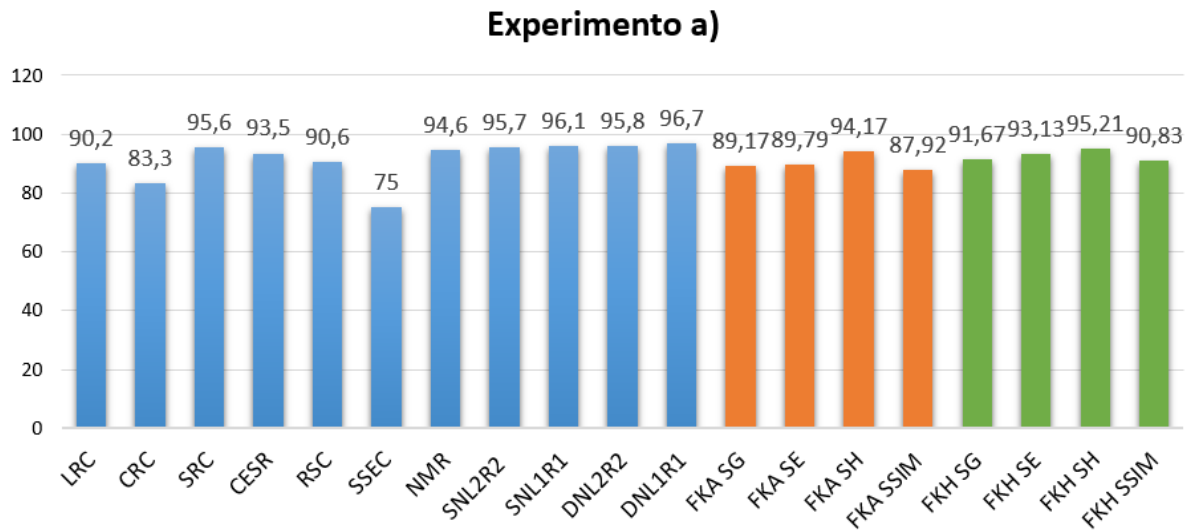


Figura 23 – Acurácias de classificação obtidas por diversos modelos da literatura no experimento a). Os resultados dos onze primeiros modelos foram extraídos de [47].

norma-1, \mathcal{S}_H e a medida de similaridade estrutural SSIM. Todos os experimentos foram realizados no *software* MATLAB. Calculamos a medida SSIM utilizando o algoritmo disponível em [2], com os parâmetros $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, $K_1 = 0,01$, $K_2 = 0,03$ conforme sugerido em [2, 79] e $L = 1$.

Denotando por FKA os classificadores baseados nas abordagens autoassociativa e por e FKH os classificadores baseados na abordagem heteroassociativa da *fuzzy*-KAM, apresentamos nas Figuras 23 e 24 os gráficos com os resultados obtidos nos experimentos a) e b) respectivamente. Com o objetivo de comparar os resultados, os gráficos apresentados nas Figuras 23 e 24 também mostram os resultados obtidos por outros modelos da literatura nos mesmos experimentos. Os resultados dos onze primeiros modelos foram extraídos de [47].

De acordo com o gráfico apresentado na Figura 23, no experimento a), os classificadores baseados em ambas as abordagens autoassociativa e heteroassociativa, com as medidas de similaridade consideradas, obtiveram resultados razoáveis, próximos dos obtidos por alguns dos modelos da literatura representados pelas barras azuis no gráfico da Figura 23. No entanto, conforme podemos observar no gráfico da Figura 24, os classificadores baseados na *fuzzy*-KAM com a medida de similaridade estrutural (SSIM) obtiveram um ótimo desempenho no experimento b), superando os resultados obtidos pelos outros modelos. De modo geral, observamos também que, em ambos os experimentos, os classificador baseado em ambas as abordagens propostas obtiveram desempenhos semelhantes.

Fizemos também um estudo do classificador heteroassociativo baseado na *fuzzy*-

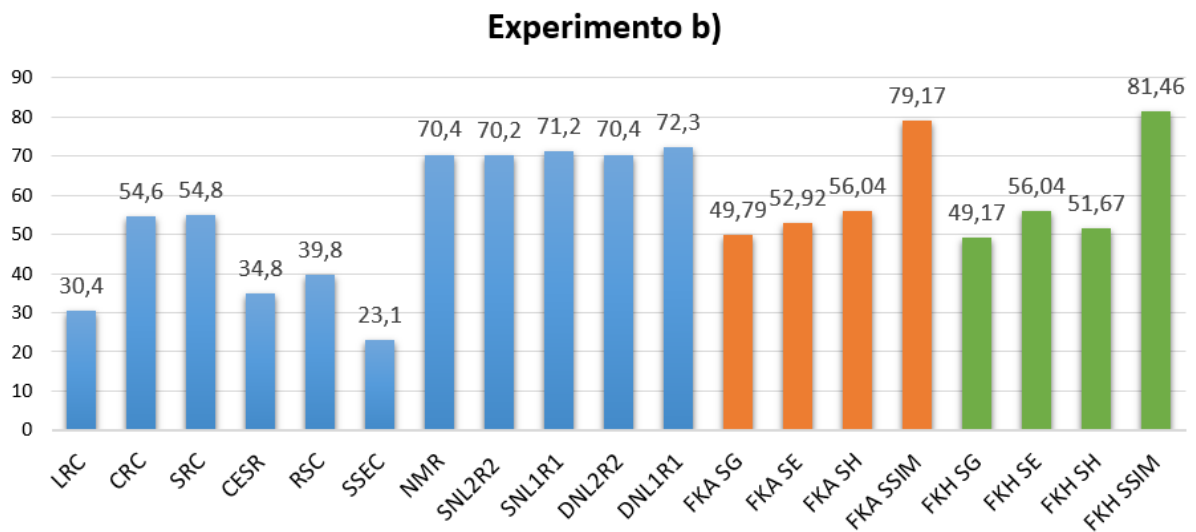


Figura 24 – Acurácias de classificação obtidas por diversos modelos da literatura no experimento b). Os resultados dos onze primeiros modelos foram extraídos de [47].

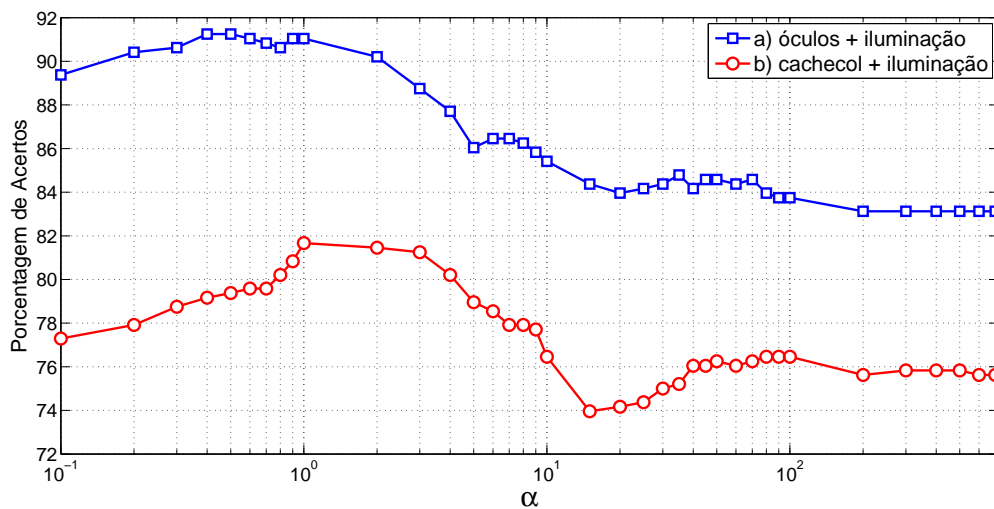


Figura 25 – Acurácias de classificação obtidas pelo classificador heteroassociativo baseado na *fuzzy*-KAM pelo valor do parâmetro α considerando imagens dos indivíduos com a) óculos+iluminação e b) cachecol+iluminação. Gráfico extraído de [69].

KAM com a medida de similaridade SSIM considerando diversos valores do parâmetro α , o qual foi apresentado em [69]. As porcentagens de acerto obtidas pelo classificador baseado na abordagem heteroassociativa nos experimentos a) e b) para vários valores do parâmetro α são mostrados na Figura 25.

No experimento a), o melhor resultado foi uma porcentagem de acerto de 91,25%, obtido para $\alpha = 0,4$ e $\alpha = 0,5$. No experimento b), o classificador heteroassociativo com a medida SSIM atingiu uma porcentagem de acertos de 81,67% com $\alpha = 1$. Note que esses

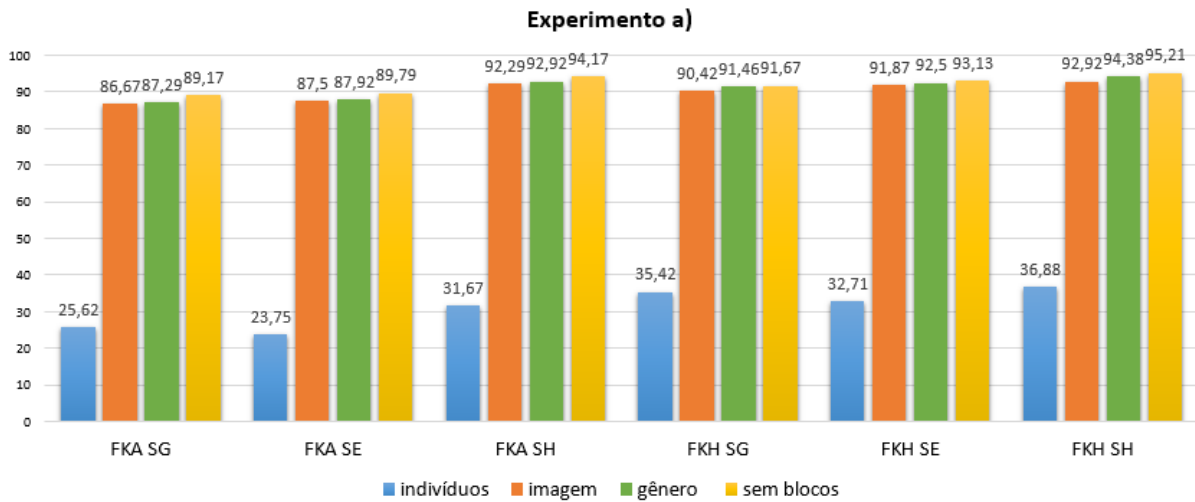


Figura 26 – Acurácias de classificação obtidas por classificadores baseados em ambas as abordagens propostas, no experimento a), considerando a matriz G com estrutura em blocos.

valores não são muito diferentes daqueles obtidos considerando α que maximiza a entropia. Sobretudo, mesmo no pior caso, tal classificador obteve uma taxa de acerto de 74% no experimento b), a qual supera os resultados obtidos pelos modelos mostrados no gráfico da Figura 24.

Experimentos utilizando a matriz G com estrutura em blocos

Utilizando a formulação matricial da *fuzzy*-KAM descrita no Capítulo 5, sugerimos uma alternativa para diminuir o custo computacional. Especificamente, em vez de considerar a matriz G sugerida pelo Teorema 3, utilizamos uma matriz G com uma estrutura em blocos esparsa, conforme (5.37). Tal alternativa se mostrou útil no caso em que o número de blocos foi definido de forma adequada usando informações adicionais sobre os indivíduos, como por exemplo gênero, expressão facial e condições de iluminação. Realizamos novamente os experimentos a) e b) com a base de imagens faciais AR descritos anteriormente e consideramos três medidas de similaridade, a saber: a medida de similaridade de Gregson, denotada por \mathcal{S}_G , a medida de similaridade de Eisler e Ekman, \mathcal{S}_E e a medida de similaridade baseada na norma-1, \mathcal{S}_H . Com cada uma dessas medidas, realizamos três experimentos com diferentes formas de construir os blocos da matriz G . Em todos os experimentos, consideramos o parâmetro α que maximiza a entropia.

As Figuras 26 e 27 mostram, respectivamente, os gráficos com os resultados obtidos por ambos os classificadores baseados na *fuzzy*-KAM nos experimentos a) e b) para diferentes formas de construir os blocos da matriz G , as quais serão descritas a seguir. Nas Figuras 26 e 27, FKA e FKH denotam, respectivamente, os classificadores baseados nas abordagens autoassociativa e heteroassociativa da *fuzzy*-KAM.

No primeiro experimento, os blocos foram divididos por indivíduo, isto é,

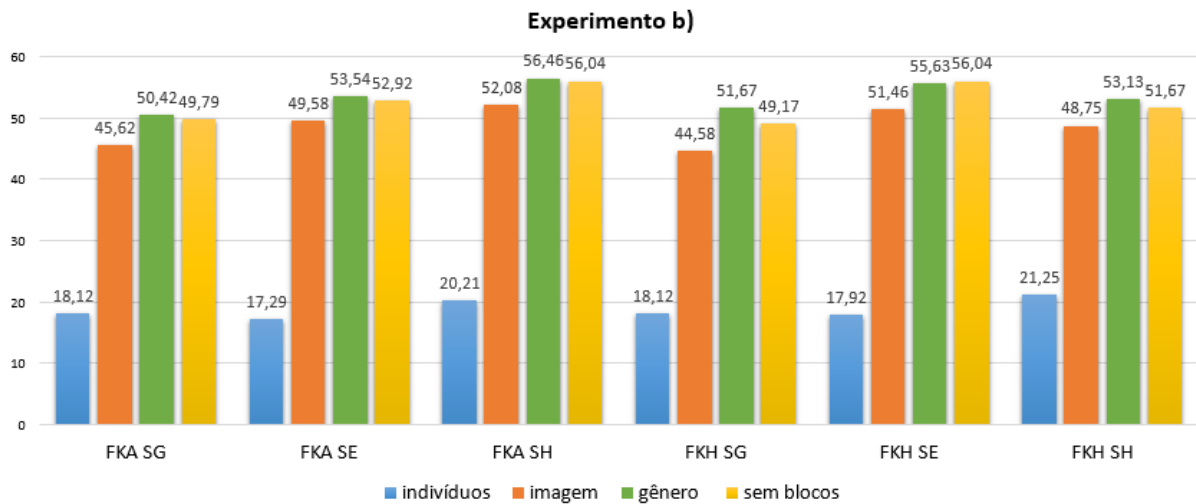


Figura 27 – Acurácias de classificação obtidas por classificadores baseados em ambas as abordagens propostas, no experimento b), considerando a matriz G com estrutura em blocos.

consideramos uma matriz G com 120 blocos (não nulos) de dimensão 8×8 em torno da diagonal principal. Os resultados desse experimento estão representados pelas colunas em azul nos gráficos das Figuras 26 e 27. Como podemos observar, essa divisão não obteve resultados satisfatórios. Isto se deve ao fato de que, considerando apenas blocos por indivíduo, deixamos de considerar a relações entre os indivíduos, o que prejudicou o desempenho do modelo.

No segundo experimento, dividimos os blocos com base na expressão facial e iluminação. De fato, a base de dados AR apresenta imagens dos indivíduos com diferentes expressões faciais e condições de iluminação. Assim, construímos um bloco utilizando a primeira imagem de cada indivíduo, um bloco com a segunda imagem de cada indivíduo e assim por diante. Dessa forma, obtivemos uma matriz G com 8 blocos (não nulos) de dimensão 120×120 em torno da diagonal principal. Os resultados desse experimento, representados pelas colunas em laranja nos gráficos das Figuras 26 e 27, mostram um bom desempenho de ambos os classificadores considerando essa divisão. Isto sugere que os blocos devem ser divididos levando em consideração, além de características comuns entre as imagens, a relação entre imagens das diferentes classes.

Por fim, no terceiro experimento, a divisão foi feita com base no gênero dos indivíduos. A base de dados AR apresenta imagens de 65 homens e 55 mulheres. Assim, consideramos apenas dois blocos. O primeiro, de dimensão 520×520 , foi formado com base nas 8 imagens de cada um dos 65 homens. O segundo, de dimensão 440×440 , foi formado considerando as 8 imagens de cada uma das 55 mulheres. Com base os resultados desse experimento, mostrado pelas colunas em verde nos gráficos das Figuras 26 e 27, podemos concluir pelo desempenho dos classificadores que essa divisão também foi adequada.



Figura 28 – Imagens de dois indivíduos da base de dados ORL.

Para efeitos de comparação, os gráficos das Figuras 26 e 27 também mostram nas colunas em amarelo, os resultados obtidos considerando a matriz G sugerida pelo Teorema 3 e portanto, sem estrutura esparsa em blocos.

2. Experimentos com a Base de Imagens Faciais ORL

A base ORL é uma base de imagens faciais do laboratório AT&T Laboratories Cambridge [60]. Esta base é composta por 400 imagens em tons de cinza com 112x92 pixels, sendo 10 imagens de cada indivíduo de um grupo de 40 pessoas, com diferentes expressões faciais. Em todas as imagens, os indivíduos estão em posição frontal, com certa tolerância para movimentos laterais e o fundo é escuro e homogêneo. Ainda, nas imagens de alguns indivíduos há variação nas condições de iluminação e em detalhes faciais como por exemplo, presença ou ausência de óculos [60]. A Figura 28 mostra imagens de dois indivíduos da base de dados ORL. Note que um deles aparece com óculos em algumas imagens e sem óculos em outras.

Tal como descrito em [20], os testes foram realizados utilizando as N primeiras imagens para formar o conjunto de treinamento e as $10 - N$ restantes como imagens de teste. Este processo foi feito para $N \in \{3, 4, 5, 6, 7\}$. Consideramos o parâmetro α que maximiza a entropia e as medidas de Gregson, \mathcal{S}_G , de Eisler e Ekman, \mathcal{S}_E , a medida de

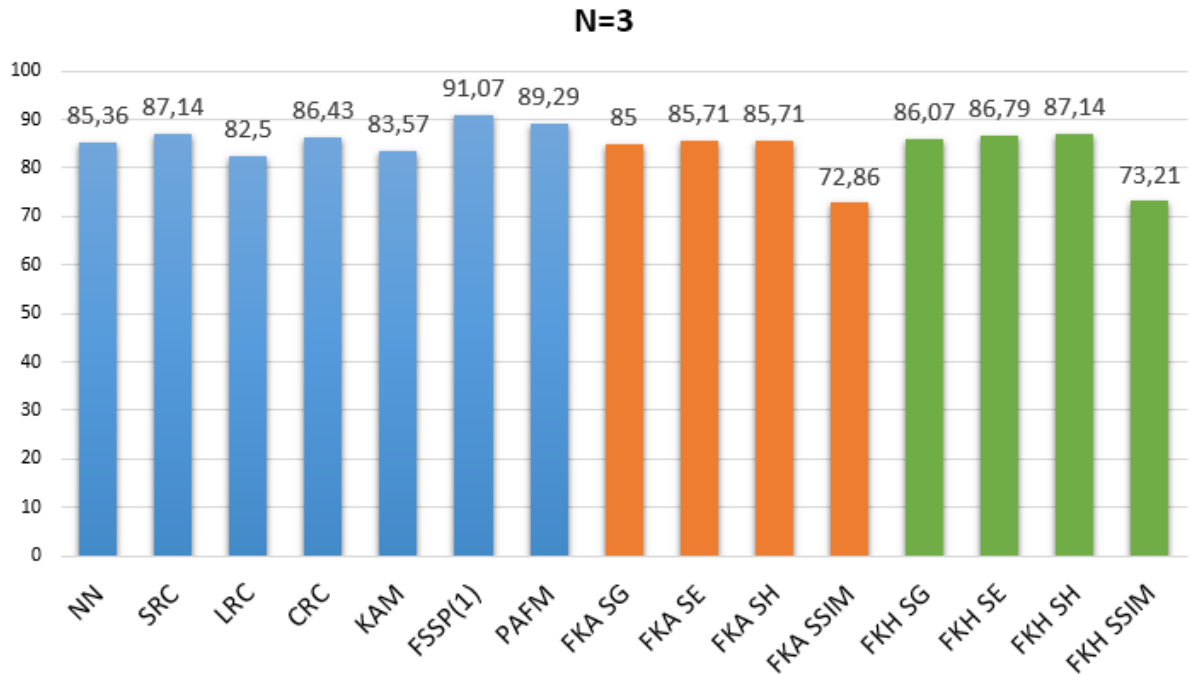


Figura 29 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais ORL, considerando $N=3$. Os resultados dos sete primeiros modelos foram extraídos de [61].

similaridade baseada na norma-1, \mathcal{S}_H , e a medida de similaridade estrutural SSIM.

Os resultados desses experimentos para $N = 3$, $N = 4$, $N = 5$, $N = 6$ e $N = 7$ são apresentados nos gráficos das Figuras 29, 30, 31, 32 e 33 respectivamente. Por motivos de comparação, estes gráficos apresentam, além da acurácia de classificação obtida pelos classificadores baseados nas *fuzzy*-KAMs, a acurácia obtida pela KAM [84] e pelos seguintes modelos da literatura: NN [11], SRC [80], LRC [54], CRC [87], FSSP(1) [20] e \mathcal{V}_Z^M (PAFM de Zadeh com mascaramento) [61]. Os resultados dos sete primeiros modelos foram extraídos de [61].

Observando as gráficos das Figuras 29, 30, 31, 32 e 33, podemos concluir que os classificadores baseados na *fuzzy*-KAM com as medidas de similaridade \mathcal{S}_G , \mathcal{S}_E e \mathcal{S}_H obtiveram desempenho competitivo com os resultados obtidos pelos outros modelos da literatura, para o caso $N = 3$, isto é, quando as três primeiras imagens de cada pessoa foram utilizadas para formar o conjunto de treinamento as 7 restantes como imagens de teste. Para o caso $N = 4$, o desempenho obtido foi razoável. No entanto, nos outros casos, os classificadores baseados na *fuzzy*-KAM com as medidas de similaridade \mathcal{S}_G , \mathcal{S}_E e \mathcal{S}_H bem como os classificadores baseados na *fuzzy*-KAM com a medida de similaridade estrutural não obtiveram bom desempenho.

3. Experimentos com a Base de Imagens Faciais GT

A base de imagens GT é formada por 750 imagens, sendo 15 imagens de cada

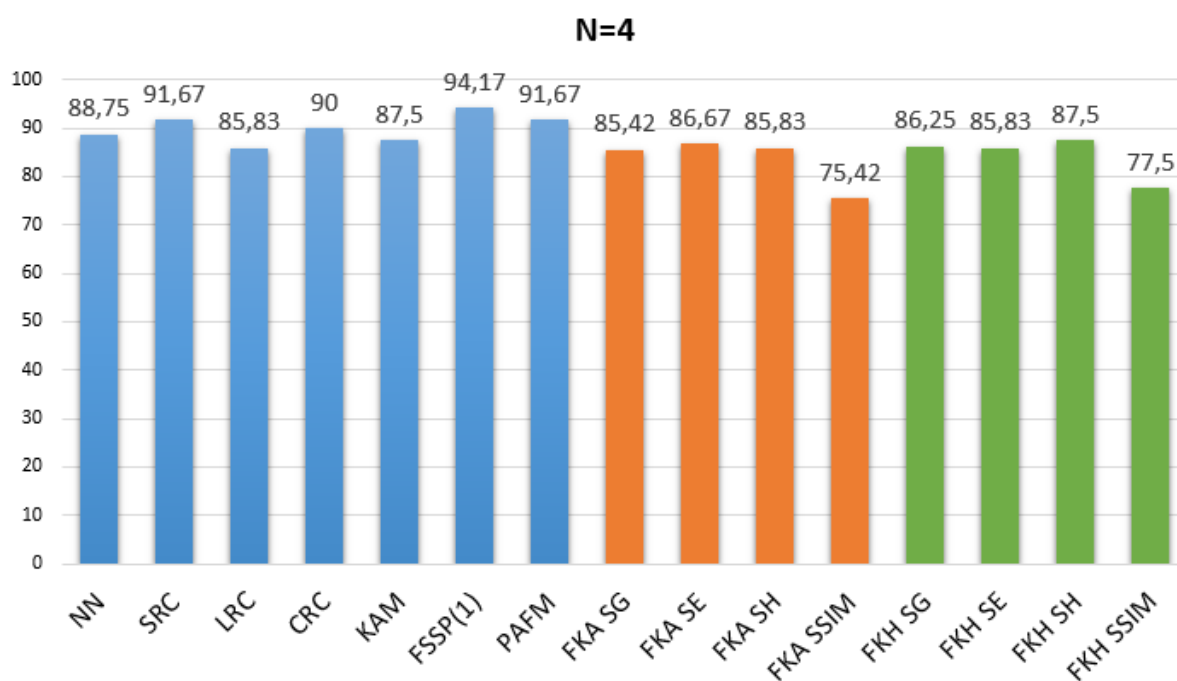


Figura 30 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais ORL, considerando $N=4$. Os resultados dos sete primeiros modelos foram extraídos de [61].

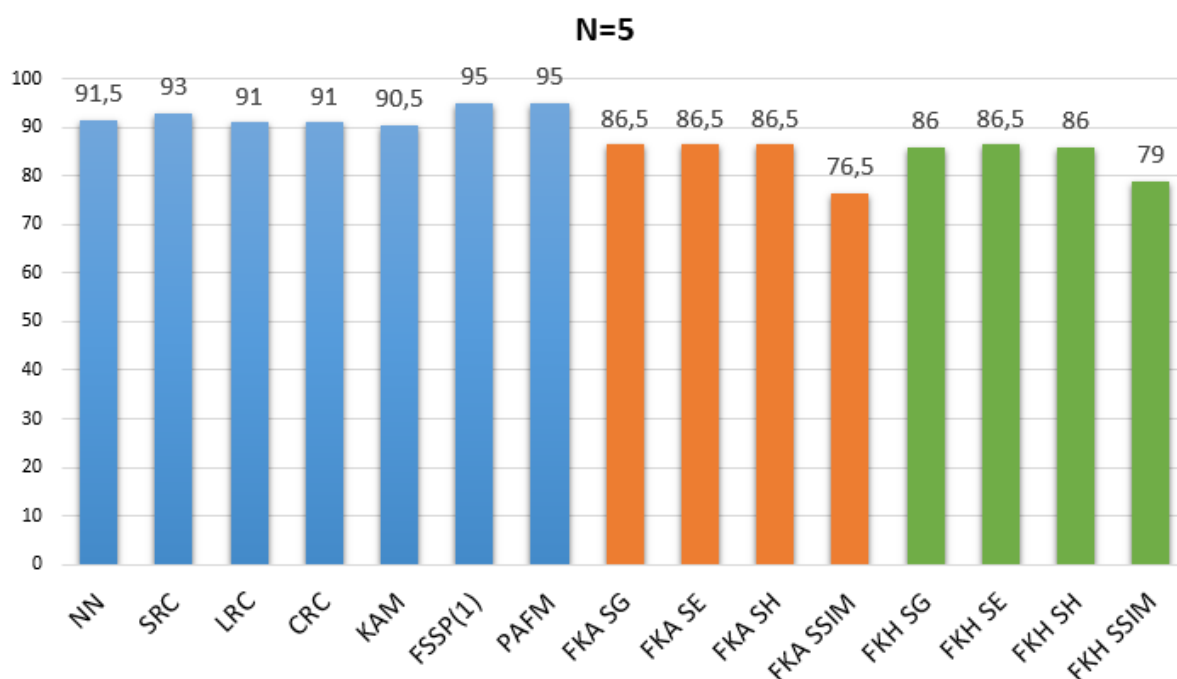


Figura 31 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais ORL, considerando $N=5$. Os resultados dos sete primeiros modelos foram extraídos de [61].

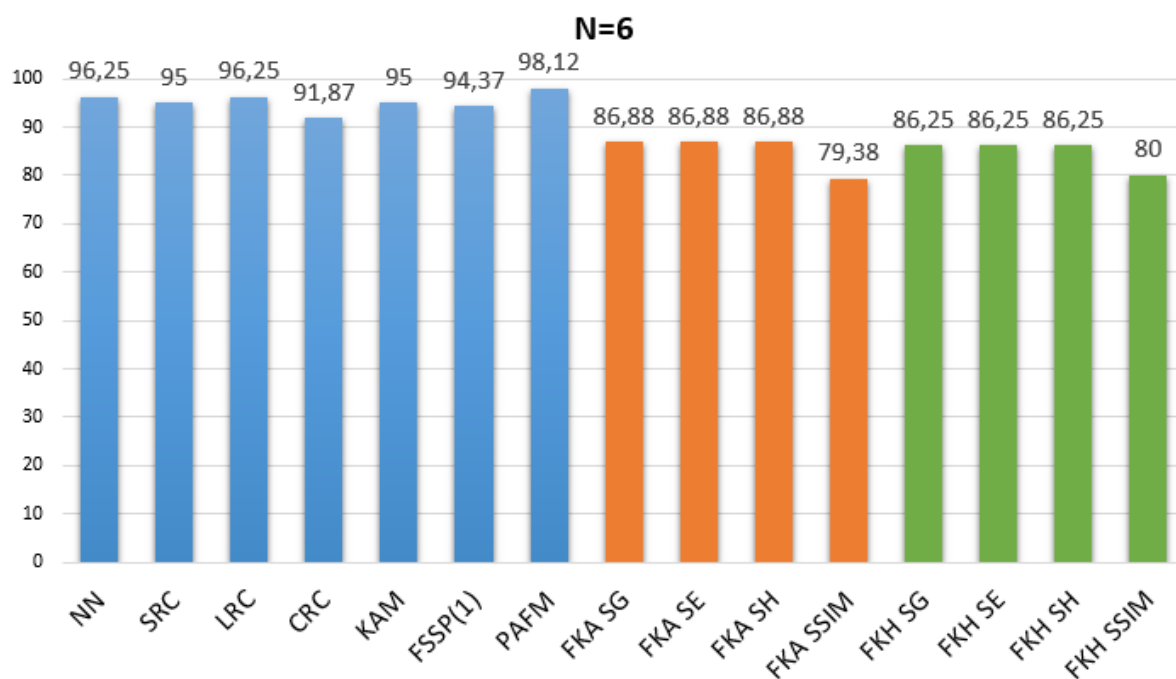


Figura 32 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais ORL, considerando N=6. Os resultados dos sete primeiros modelos foram extraídos de [61].

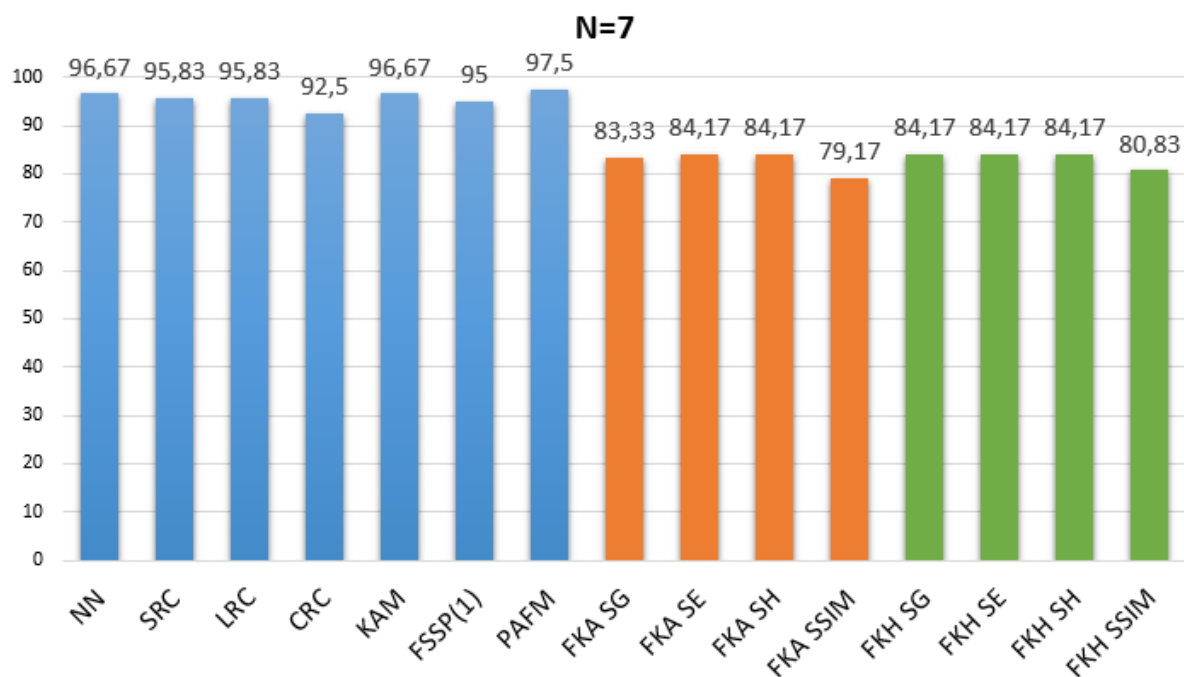


Figura 33 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais ORL, considerando N=7. Os resultados dos sete primeiros modelos foram extraídos de [61].

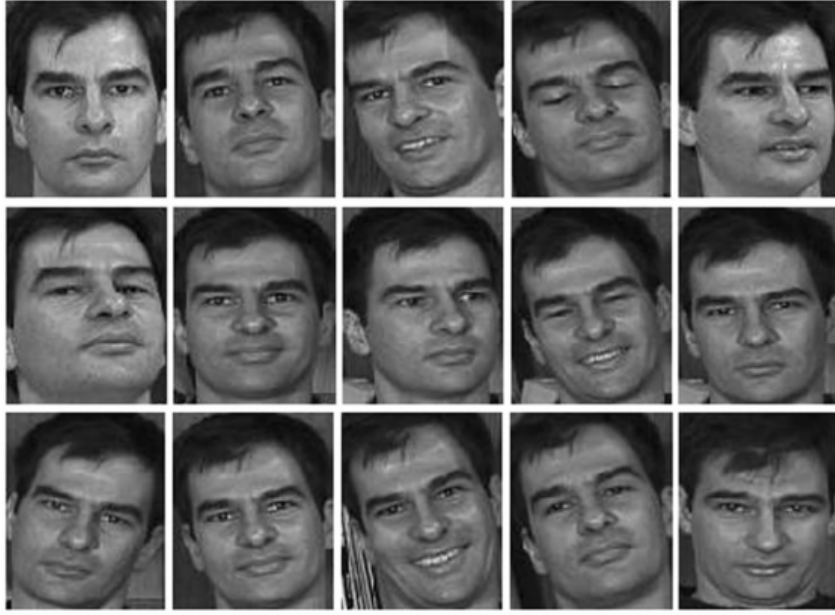


Figura 34 – Imagens de um indivíduo da base de imagens faciais GT.

indivíduo de um grupo de 50 pessoas. As imagens foram obtidas em duas ou três sessões no Centro de Processamento de Sinal e Imagem no Instituto de Tecnologia da Geórgia (*Center for Signal and Image Processing at Georgia Institute of Technology*) [1]. As imagens são coloridas, apresentam fundo desordenado e resolução 640x480 pixels. O tamanho médio das faces nas imagens é 150×150 pixels. No entanto, para os experimentos, cada imagem foi transformada em uma imagem em tons de cinza de dimensão 40×30 . A Figura 34 mostra as imagens transformadas de um indivíduo pertencentes a base de imagens GT.

De modo semelhante ao experimento realizado com a base de imagens ORL, utilizamos as N primeiras imagens para compor o conjunto de treinamento e as $15 - N$ restantes para o conjunto de teste, com $N \in \{3, 4, 5, 6, 9\}$. Também consideramos o parâmetro α que maximiza a entropia e as medidas de Gregson, \mathcal{S}_G , de Eisler e Ekman, \mathcal{S}_E , a medida de similaridade baseada na norma-1, \mathcal{S}_H , e a medida de similaridade estrutural SSIM. Os resultados desse experimento para $N = 3$, $N = 4$, $N = 5$, $N = 6$ e $N = 9$ são mostrados pelos gráficos das Figuras 35, 36, 37, 38 e 39, respectivamente. Estes gráficos apresentam, além das taxas de reconhecimento obtidas por ambos os classificadores baseados na *fuzzy*-KAM e pela KAM, as taxas de reconhecimento obtidas pelos seguintes modelos: NN [11], SRC [80], LRC [54], CRC [87], FSSP(1) [20] e \mathcal{V}_Z^M (PAFM de Zadeh com mascaramento) [61]. Os resultados dos sete primeiros modelos foram extraídos de [61].

Os gráficos das Figuras 35, 36, 37, 38 e 39 mostram que ambos os classificadores baseados na *fuzzy*-KAM, com exceção dos definidos com base na medida de similaridade estrutural SSIM, obtiveram um ótimo desempenho no problema de reconhecimento de faces considerando a base GT. De fato, para todos os valores de N considerados, exceto

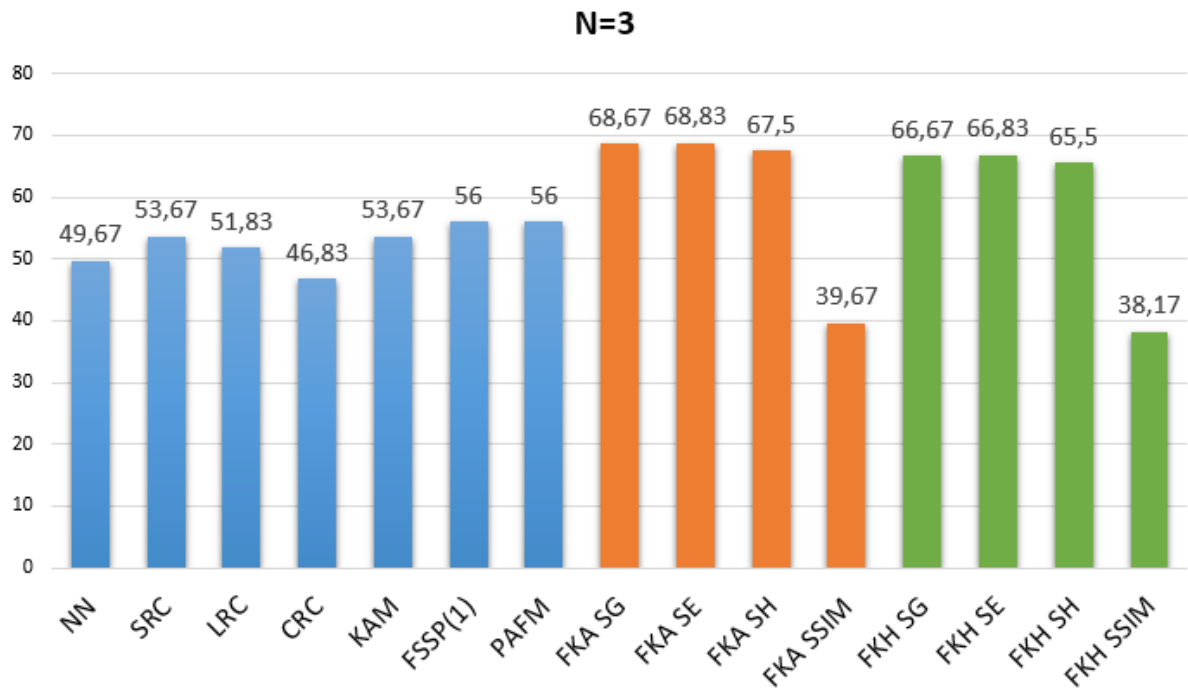


Figura 35 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais GT, considerando $N=3$. Os resultados dos sete primeiros modelos foram extraídos de [61].

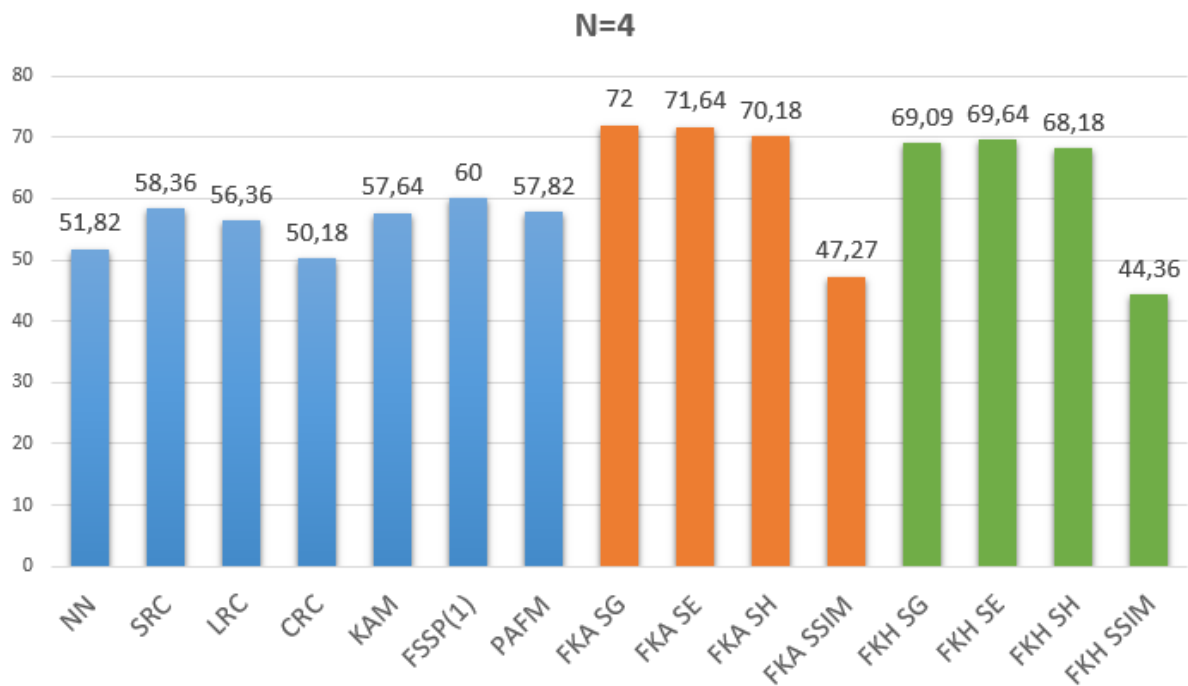


Figura 36 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais GT, considerando $N=4$. Os resultados dos sete primeiros modelos foram extraídos de [61].

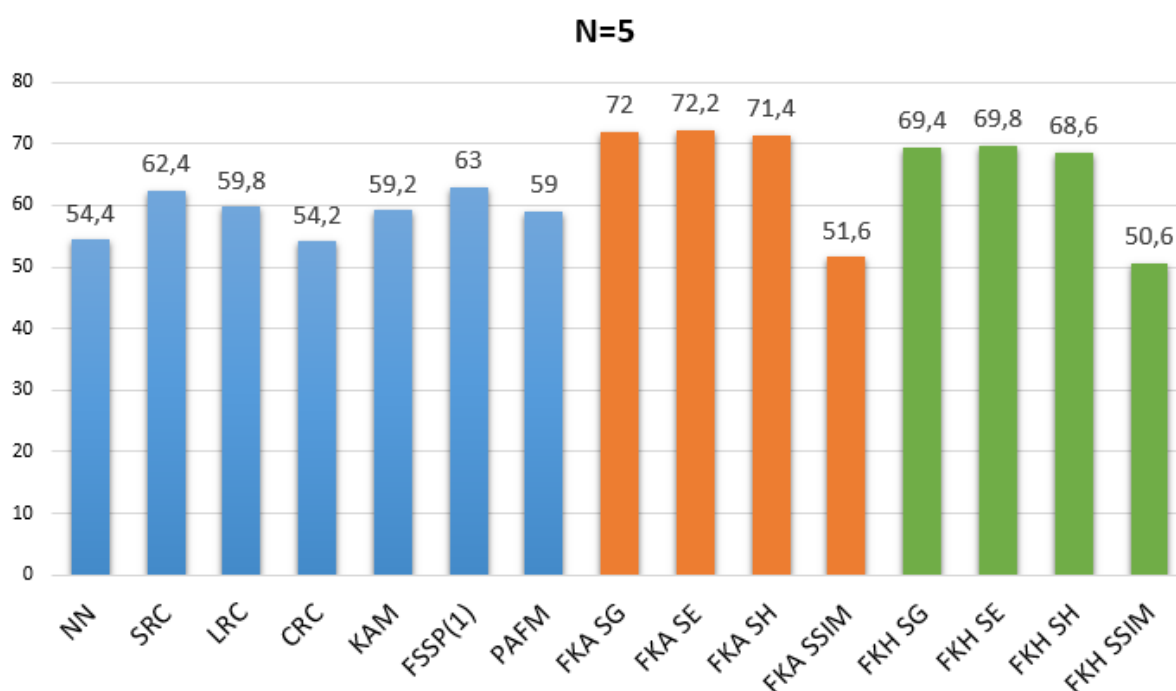


Figura 37 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais GT, considerando N=5. Os resultados dos sete primeiros modelos foram extraídos de [61].

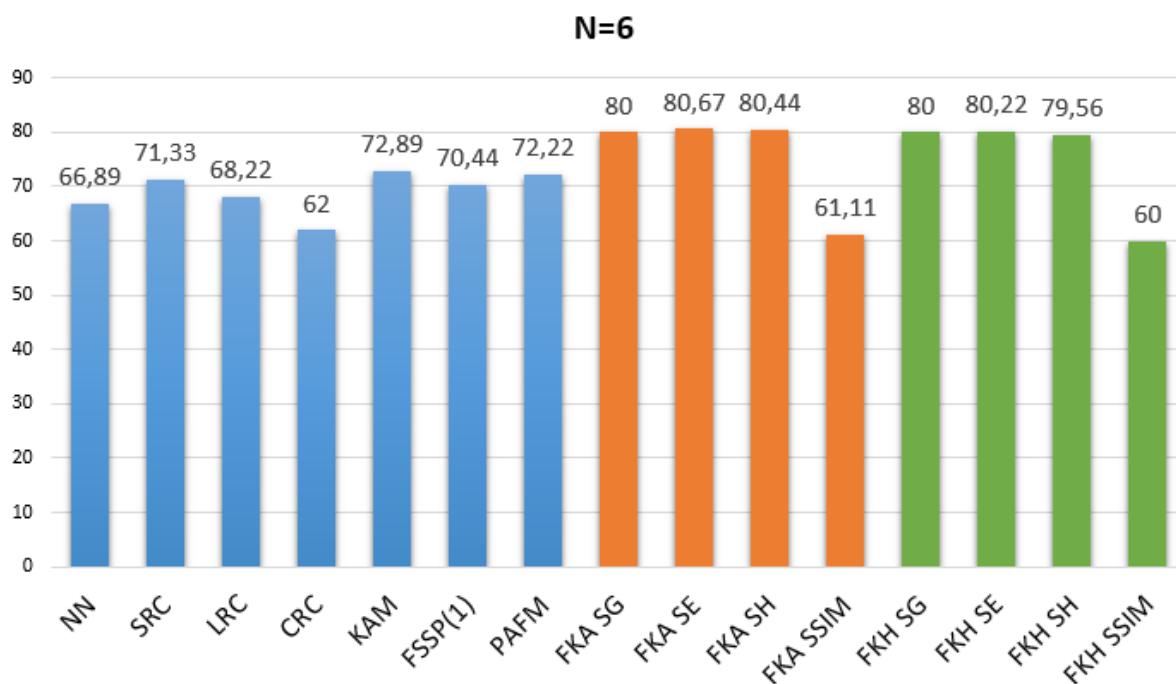


Figura 38 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais GT, considerando N=6. Os resultados dos sete primeiros modelos foram extraídos de [61].

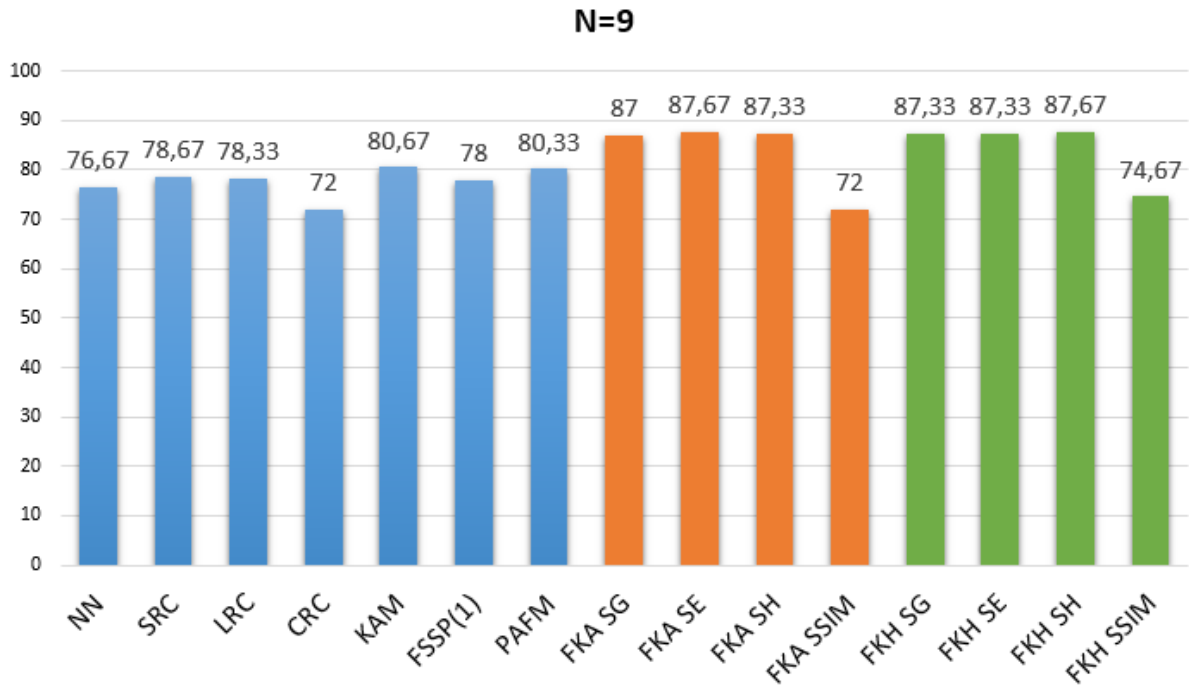


Figura 39 – Acurácia de classificação obtida por diversos modelos no experimento com a base de imagens faciais GT, considerando $N=9$. Os resultados dos sete primeiros modelos foram extraídos de [61].

pelos definidos com base na medida de similaridade estrutural SSIM, os classificadores baseados na *fuzzy-KAM* superaram os resultados produzidos pelos outros modelos da literatura.

Apesar do excelente desempenho dos classificadores baseados na *fuzzy-KAM* com a medida de similaridade SSIM no experimento b) com a base de dados AR, como observamos na Figura 24, tais classificadores com a medida SSIM não obtiveram bons resultados nos experimentos com as bases ORL e GT. Isto pode ter ocorrido pelo fato das imagens das bases ORL e GT possuírem fundo escuro enquanto que as imagens da base AR possuem fundo branco, além de que, nas imagens da base AR usadas pertencentes ao conjunto de teste do experimento b), parte das faces são ocultas pelo cachecol, enquanto que nas imagens das bases GT e ORL isto não ocorre.

Resumindo, nos problemas de classificação, os valores de acurácia obtidos foram competitivos ou superaram os obtidos por outros modelos da literatura. Nos problemas de reconhecimento de faces, apesar de não terem obtido bom desempenho nos experimentos com a base GT utilizando a medida de similaridade estrutural e nos experimentos com a base de imagens ORL, os classificadores baseados em ambas as abordagens propostas obtiveram excelentes resultados nos experimentos com a base de imagens GT com as medidas \mathcal{S}_G , \mathcal{S}_E e \mathcal{S}_H bem como no experimento b) com a base de dados AR utilizando a medida de similaridade estrutural SSIM.

Dessa forma, de modo geral, podemos concluir que, enquanto outros modelos da literatura são desenvolvidos especificamente para problemas de classificação ou problemas de reconhecimento de faces, os classificadores baseados tanto na abordagem autoassociativa como na heteroassociativa apresentaram resultados satisfatórios em ambos os tipos de problema, o que mostra a versatilidade das abordagens propostas.

7 Considerações Finais

Nesta tese, inicialmente, fizemos uma revisão dos principais modelos da literatura que motivaram o desenvolvimento do trabalho. Especificamente, recordamos as definições de memórias associativas para armazenamento e recordação de vetores bipolares, a saber: a rede de Hopfield [34], as memórias associativas recorrentes por correlação (RCAMs) [9], as memórias associativas exponenciais por correlação (ECAMs) [9] e as memórias associativas exponenciais bidirecionais (EBAMs) [37]. Revisamos também as memórias associativas recorrentes exponenciais multivaloradas [10] (MERAMs) e a memória associativa com núcleo (KAMs) [84], modelos que podem ser vistos como uma extensão das RCAMs para o armazenamento e recordação de vetores reais. Em seguida, recordamos alguns modelos de memórias associativas *fuzzy*. Iniciamos com a definição das memórias associativas recorrentes exponenciais *fuzzy* (RE-FAMs) [75], as quais podem ser interpretadas como uma versão *fuzzy* das ECAMs [9]. Posteriormente, apresentamos a definição das memórias associativas recorrentes exponenciais *fuzzy* generalizadas (GRE-FAMs) [64, 70], modelos que generalizam as RE-FAMs por meio do acréscimo de uma camada oculta de neurônios lineares com o objetivo de reduzir a interferência cruzada entre as memórias fundamentais. Finalmente, apresentamos as Θ -FAMs [17], memórias associativas *fuzzy* que apresentam certas relações com as GRE-FAMs.

Como um dos principais objetivos, apresentamos uma extensão das GRE-FAMs para o caso heteroassociativo, modelo denominado GEB-FAM. Mostramos que a GEB-FAM possui capacidade ótima de armazenamento quando consideramos a matriz adequada em sua definição e apresentamos uma caracterização teórica do primeiro passo da GEB-FAM quando seu parâmetro α tende a infinito. No entanto, a dinâmica do modelo ainda não está bem compreendida. De fato, por meio de um exemplo, observamos que a GEB-FAM nem sempre converge se o parâmetro α não for suficientemente grande. Apesar disso, valores menores do parâmetro α (geralmente entre 1 e 10) produziram melhores resultados na maioria dos experimentos computacionais. Por este motivo, investigamos o primeiro passo da GEB-FAM. Tal modelo, denominado *fuzzy*-KAM, foi abordado utilizando núcleos *fuzzy*.

Apresentamos reformulações dos resultados teóricos apresentados para a GEB-FAM para o caso das *fuzzy*-KAM. Além disso, definimos uma condição suficiente para o parâmetro α a partir do qual a matriz K , necessária na definição da *fuzzy*-KAM, é definida positiva e propusemos o ajuste deste parâmetro utilizando o conceito de entropia. Precisamente, definimos a entropia de uma *fuzzy*-KAM com base na entropia de Shannon e propusemos utilizar o valor de α que maximiza a entropia. Por fim, apresentamos uma formulação matricial da *fuzzy*-KAM e sugerimos uma forma de definir a matriz $G = K^{-1}$ esparsa com estrutura em blocos, alternativa que pode ser útil, sobretudo nos casos em

que temos um grande número de memórias fundamentais.

Também fizemos um estudo de várias medidas de similaridade da literatura. Consideramos inicialmente uma definição abrangente proposta por De Baets et. al. Nesse contexto, estudamos as medidas de similaridade baseadas em cardinalidade e uma versão normalizada da medida de similaridade proposta por Xuecheng. Também consideramos a medida de similaridade estrutural SSIM, a qual apesar de não ser uma medida de similaridade *fuzzy*, pode ser empregada para definir um núcleo *fuzzy* κ , para imagens em tons de cinza.

Como aplicações, estudamos problemas de classificação e reconhecimento de faces. Especificamente, apresentamos duas abordagens para classificação de padrões utilizando *fuzzy*-KAMs. A primeira abordagem, baseada em classificadores de representação esparsa, é definida utilizando uma *fuzzy*-KAM autoassociativa, enquanto que a segunda, se baseia em uma abordagem heteroassociativa. Experimentos computacionais foram realizados para avaliar o desempenho das abordagens propostas. Consideramos diversos problemas de classificação disponíveis nos repositórios KEEL e UCI, além de problemas de reconhecimento de faces considerando as bases de imagens faciais AR, ORL e GT. Avaliamos também o desempenho dos modelos com as diferentes medidas de similaridade descritas no Capítulo 2.

Nos problemas de classificação os resultados obtidos foram satisfatórios. De fato, os classificadores baseados na *fuzzy*-KAM, definidos com base em ambas as abordagens propostas, obtiveram, na maioria dos problemas, resultados competitivos ou que superaram os obtidos por outros modelos da literatura.

Nos problemas de reconhecimento de faces os classificadores baseados na *fuzzy*-KAM também obtiveram bons resultados. De forma geral, considerando a base de imagens faciais AR, no experimento a), os classificadores heteroassociativos baseados na *fuzzy*-KAM apresentaram melhor desempenho que os baseados na abordagem autoassociativa e produziram taxas de acertos competitivas com as de outros modelos da literatura. No mesmo experimento, os classificadores baseados na versão autoassociativa da *fuzzy*-KAM apresentaram resultados razoáveis, próximos dos resultados obtidos por alguns modelos da literatura. No entanto, no experimento b), ambos os classificadores baseados na *fuzzy*-KAM com a medida de similaridade estrutural SSIM apresentaram excelente desempenho, superando os resultados dos outros modelos da literatura considerados. Para o classificador heteroassociativo com as medidas de similaridade baseadas em cardinalidade, os melhores resultados foram obtidos pelas medidas *T*-transitivas definidas usando a *t*-norma do mínimo nos esquemas de fuzzificação. Este resultado sugere que além da *T*-transitividade, a reflexividade é uma propriedade importante a ser considerada. As medidas *T*-transitivas definidas usando a *t*-norma do produto também produziram bons resultados.

Os classificadores baseados na *fuzzy*-KAM não obtiveram bons resultados para a

base ORL. De fato, tais classificadores obtiveram resultado competitivo, se comparados com outros classificadores, apenas no caso em que o conjunto de treinamento foi formado por três imagens de cada indivíduo. No caso em que consideramos um conjunto de treinamento formado por quatro imagens de cada indivíduos, os resultados obtidos foram razoáveis. No entanto, considerando a base de dados GT, os classificadores baseados na *fuzzy-KAM* com as medidas de similaridade de Gregson, de Eisler e Ekman e a medida baseada na norma-1 superaram os resultados dos outros modelos da literatura em todos os casos.

Concluindo, apesar de não terem obtido bom desempenho em todos os experimentos realizados, de modo geral, ambos os classificadores heteroassociativo e autoassociativo baseados nas *fuzzy-KAM* apresentaram resultados satisfatórios tanto para problemas de classificação como para problemas de reconhecimentos de faces, o que mostra a versatilidade das abordagens propostas.

Como trabalhos futuros, podemos investigar estratégias de seleção de memórias fundamentais com o objetivo de diminuir o custo computacional exigido na fase de armazenamento da *fuzzy-KAM*. Ainda com este objetivo, considerando a proposta apresentada de definir a matriz G (presente na definição da *fuzzy-KAM*) com estrutura esparsa em blocos, é necessário estabelecer critérios para definir de forma eficiente a quantidade de blocos e como estes serão definidos. Além disso, um estudo mais aprofundado dos núcleos *fuzzy*, sua representatividade e interpretação, bem como das propriedades da função entropia da *fuzzy-KAM* pode ser realizado. Por fim, pode-se explorar outras aplicações da *fuzzy-KAM*.

Referências

- [1] Georgia Tech Face Database. Disponível em http://www.anefian.com/research/face_reco.htm Acesso em Fevereiro, 2018. Citado 4 vezes nas páginas 18, 76, 80 e 93.
- [2] The SSIM Index for Image Quality Assessment. Disponível em: <http://ece.uwaterloo.ca/~z70wang/research/ssim/> Acesso em Fevereiro, 2017. Citado na página 85.
- [3] J. Alcalá-Fdez, R. Alcalá, and F. Herrera. A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems*, 19(5):857–872, 2011. Citado 4 vezes nas páginas 76, 77, 78 e 79.
- [4] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(2-3):255–287, 2011. Citado 2 vezes nas páginas 18 e 76.
- [5] J. Anderson. A Simple Neural Network Generating Interactive Memory. *Mathematical Biosciences*, 14:197–220, 1972. Citado na página 35.
- [6] K. Bache and M. Lichman. UCI machine learning repository, 2013. Citado 2 vezes nas páginas 18 e 76.
- [7] L. Barros and R. Bassanezi. *Tópicos de Lógica Fuzzy e Biomatemática*. Editora do Instituto de Matemática, Estatística e Computação Científica (IMECC-Unicamp), Campinas, 2006. Citado na página 21.
- [8] B. Bede. *Mathematics of Fuzzy Sets and Fuzzy Logic*, volume 295 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag Berlin Heidelberg, 2013. Citado 2 vezes nas páginas 19 e 21.
- [9] T. D. Chiueh and R. M. Goodman. Recurrent Correlation Associative Memories. *IEEE Transactions on Neural Networks*, 2(2):275–284, 1991. DOI:10.1109/72.80338. Citado 8 vezes nas páginas 12, 17, 28, 30, 31, 32, 33 e 98.
- [10] T.-D. Chiueh and H.-K. Tsai. Multivalued associative memories based on recurrent networks. *IEEE Transactions on Neural Networks*, 4(2):364–366, March 1993. Citado 6 vezes nas páginas 12, 17, 28, 33, 34 e 98.

- [11] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, January 1967. Citado 2 vezes nas páginas 90 e 93.
- [12] B. De Baets and H. De Meyer. Transitivity-preserving fuzzification schemes for cardinality-based similarity measures. *European Journal of Operational Research*, 160(3):726 – 740, 2005. DOI:10.1016/j.ejor.2003.06.036. Citado 10 vezes nas páginas 13, 15, 17, 21, 22, 23, 24, 26, 51 e 81.
- [13] B. De Baets, H. De Meyer, and H. Naessens. A class of rational cardinality-based similarity measures. *Journal of Computational and Applied Mathematics*, 132(1):51–69, 2001. Citado na página 23.
- [14] A. De Luca and S. Termini. A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and control*, 20(4):301–312, 1972. Citado na página 60.
- [15] R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J.-J. Schmid, S. Sandhu, K. H. Guppy, S. Lee, and V. Froelicher. International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 64(5):304–310, 1989. Citado na página 76.
- [16] D. Dubois and H. Prade. *Fuzzy sets and systems: theory and applications*. Academic Press, New York, 1980. Citado na página 25.
- [17] E. Esmi, P. Sussner, H. Bustince Sola, and J. Fernandez. Θ -fuzzy associative memories (Θ -FAMs). *IEEE Transactions on Fuzzy Systems*, 23(2):313 – 326, 2015. Citado 6 vezes nas páginas 13, 17, 28, 40, 41 e 98.
- [18] E. Esmi, P. Sussner, and S. Sandri. Tunable equivalence fuzzy associative memories. *Fuzzy Sets and Systems*, 292:242–260, 2016. Citado 3 vezes nas páginas 13, 17 e 41.
- [19] E. L. Esmi, P. Sussner, M. E. Valle, F. Sakuray, and L. Barros. Fuzzy Associative Memories Based on Subsethood and Similarity Measures with Applications to Speaker Identification. In *Lecture Notes in Computer Science: International Conference on Hybrid Artificial Intelligence Systems (HAIS 2012)*, pages 479–490. Springer-Verlag Berlin Heidelberg, Berlin, Germany, 2012. Citado na página 41.
- [20] Q. Feng, C. Yuan, J. S. Pan, J. F. Yang, Y. T. Chou, Y. Zhou, and W. Li. Superimposed sparse parameter classifiers for face recognition. *IEEE Transactions on Cybernetics*, 47(2):378–390, Feb 2017. Citado 3 vezes nas páginas 89, 90 e 93.
- [21] M. J. Frank. On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$. *Aequationes Mathematicae*, 19(1):194–226, 1979. Citado na página 21.

- [22] C. García and J. A. Moreno. The Hopfield Associative Memory Network: Improving Performance with the Kernel Trick. In *Lecture Notes in Artificial Intelligence - Proceedings of IBERAMIA 2004*, volume 3315 of *Advances in Artificial Intelligence - IBERAMIA 2004*, pages 871–880. Springer-Verlag, 2004. Citado 2 vezes nas páginas 14 e 51.
- [23] G. Golub and C. van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, MD, 3th edition, 1996. Citado 4 vezes nas páginas 39, 56, 57 e 66.
- [24] A. González and R. Pérez. Selection of relevant features in a fuzzy genetic learning algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(3):417–425, 2001. Citado na página 77.
- [25] H. L. Guidorizzi. *Um Curso de Cálculo*, volume 1. Editora LTC, 2005. Citado na página 59.
- [26] M. H. Hassoun, editor. *Associative Neural Memories: Theory and Implementation*. Oxford University Press, Oxford, U.K., 1993. Citado 3 vezes nas páginas 11, 12 e 28.
- [27] M. H. Hassoun. *Fundamentals of Artificial Neural Networks*. MIT Press, Cambridge, MA, 1995. Citado 2 vezes nas páginas 12 e 28.
- [28] M. H. Hassoun and P. B. Watta. Associative Memory Networks. In E. Fiesler and R. Beale, editors, *Handbook of Neural Computation*, pages C1.3:1–C1.3:14. Oxford University Press, 1997. Citado na página 11.
- [29] S. Haykin. *Neural Networks and Learning Machines*. Prentice-Hall, Upper Saddle River, NJ, 3rd edition edition, 2009. Citado na página 29.
- [30] D. Hebb. *The Organization of Behavior*. John Wiley & Sons, New York, 1949. Citado na página 29.
- [31] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220, 2008. Citado 3 vezes nas páginas 14, 36 e 51.
- [32] A. O. Hoori and Y. Motai. Multicolumn RBF network. *IEEE Transactions on Neural Networks and Learning Systems*, 29(4):766–778, 2018. Citado na página 66.
- [33] J. Hopfield and D. Tank. Neural computation of decisions in optimization problems. *Biological Cybernetics*, 52:141–152, 1985. Citado na página 12.
- [34] J. J. Hopfield. Neural Networks and Physical Systems with Emergent Collective Computational Abilities. *Proceedings of the National Academy of Sciences*, 79:2554–2558, Apr. 1982. Citado 5 vezes nas páginas 12, 17, 28, 29 e 98.

- [35] H. Ishibuchi, T. Yamamoto, and T. Nakashima. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(2):359–365, 2005. Citado na página 77.
- [36] P. Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579, 1901. Citado na página 25.
- [37] Y.-J. Jeng, C. C. Yeh, and T. D. Chiueh. Exponential bidirectional associative memories. *Electronics Letters*, 26(11):717–718, 1990. DOI:10.1049/el:19900468. Citado 7 vezes nas páginas 12, 17, 28, 33, 40, 42 e 98.
- [38] E. P. Klement, R. Mesiar, and E. Pap. Triangular norms, 2000. Citado na página 20.
- [39] G. J. Klir. *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-Interscience, 2005. Citado na página 58.
- [40] T. Kohonen. Correlation Matrix Memory. *IEEE Transactions on Computers*, C-21:353–359, 1972. Citado na página 35.
- [41] S.-G. Kong and B. Kosko. Adaptive Fuzzy Systems for Backing up a Truck-and-Trailer. *IEEE Transactions on Neural Networks*, 3(2):211–223, 1992. Citado na página 12.
- [42] B. Kosko. Bidirectional Associative Memories. *IEEE Transactions on Systems, Man, and Cybernetics*, 18:49–60, 1988. Citado na página 12.
- [43] B. Kosko. *Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence*. Prentice Hall, Englewood Cliffs, NJ, 1992. Citado 2 vezes nas páginas 12 e 20.
- [44] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pages 369–376. IEEE, 2001. Citado na página 77.
- [45] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. In *Proceedings of the fourth international conference on knowledge discovery and data mining*, 1998. Citado na página 77.
- [46] B. Liu, Y. Ma, and C.-K. Wong. Classification using association rules: weaknesses and enhancements. In *Data mining for scientific and engineering applications*, pages 591–605. Springer, 2001. Citado na página 77.
- [47] L. Luo, J. Yang, J. Qian, and Y. Tai. Nuclear-l1 norm joint regression for face reconstruction and recognition with mixed noise. *Pattern Recognition*, 48(12):3811 – 3824, 2015. DOI:10.1016/j.patcog.2015.06.012. Citado 3 vezes nas páginas 80, 85 e 86.

- [48] E. G. Mansoori, M. J. Zolghadri, and S. D. Katebi. Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data. *IEEE Transactions on Fuzzy Systems*, 16(4):1061–1071, 2008. Citado na página 77.
- [49] A. M. Martinez. The AR face database. *CVC Technical Report*, 24, 1998. Citado 3 vezes nas páginas 18, 76 e 80.
- [50] W. McCulloch and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bulletin of Mathematical Biophysics*, 5:115–133, 1943. Citado na página 29.
- [51] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. Venkatesh. The capacity of the Hopfield associative memory. *IEEE Transactions on Information Theory*, 1:33–45, 1987. Citado 2 vezes nas páginas 12 e 30.
- [52] C. D. Meyer. *Matrix analysis and applied linear algebra*. SIAM, 2000. Citado na página 56.
- [53] K. Nakano. Associatron: A Model of Associative Memory. *IEEE Trans. on Systems, Man, Cybernetics*, SMC-2:380–388, 1972. Citado na página 35.
- [54] I. Naseem, R. Togneri, and M. Bennamoun. Linear regression for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(11):2106–2112, Nov 2010. Citado 2 vezes nas páginas 90 e 93.
- [55] H. T. Nguyen and E. A. Walker. *A First Course in Fuzzy Logic*. Chapman & Hall/CRC, Boca Raton, 2 edition, 2000. Citado na página 21.
- [56] N. Pal and S. Pal. Higher order fuzzy entropy and hybrid entropy of a set. *Information Sciences*, 61(3):211 – 231, 1992. Citado na página 58.
- [57] W. Pedrycz and F. Gomide. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. Wiley-IEEE Press, New York, 2007. Citado 2 vezes nas páginas 20 e 21.
- [58] J. C. Principe. Information theory, machine learning, and reproducing kernel hilbert spaces. In *Information theoretic learning*, pages 1–45. Springer, 2010. Citado na página 57.
- [59] J. Quinlan. C4. 5: Programs for machine learning. 1993. Citado na página 77.
- [60] F. S. Samaria and A. C. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, pages 138–142, Dec 1994. Citado 4 vezes nas páginas 18, 76, 80 e 89.
- [61] A. S. Santos. *Memórias autoassociativas de Projeções em reticulados completos*. PhD thesis, Universidade Estadual de Campinas (UNICAMP), 2018. Citado 7 vezes nas páginas 90, 91, 92, 93, 94, 95 e 96.

- [62] C. E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 1948. Citado 2 vezes nas páginas 14 e 58.
- [63] S. Si, C.-J. Hsieh, and I. S. Dhillon. Memory efficient kernel approximation. *Journal of Machine Learning Research*, 18:1–32, 2017. Citado na página 66.
- [64] A. C. Souza. Memórias Associativas Recorrentes Exponenciais Fuzzy Baseadas em Medidas de Similaridade. Master’s thesis, Universidade Estadual de Campinas (UNICAMP), Campinas, Brasil, Abril 2015. Citado 10 vezes nas páginas 13, 17, 28, 38, 39, 41, 42, 48, 49 e 98.
- [65] A. C. Souza and M. E. Valle. Memória Associativa Bidirecional Exponencial Fuzzy Generalizada Aplicada ao Reconhecimento de Faces. In M. E. Valle, G. Dimuro, R. Santiago, and E. Esmi, editors, *Recentes Avanços em Sistemas Fuzzy*, volume 1, pages 503–514. Sociedade Brasileira de Matemática Aplicada e Computacional (SB-MAC), São Carlos - SP, 2016. ISBN: 978-85-8215-079-5. Citado 3 vezes nas páginas 16, 52 e 68.
- [66] A. C. Souza and M. E. Valle. Memórias associativas recorrentes exponenciais fuzzy generalizadas aplicadas à classificação de padrões. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 5(1), 2017. Citado 2 vezes nas páginas 16 e 68.
- [67] A. C. Souza and M. E. Valle. Fuzzy kernel associative memories with application in classification. In G. A. Barreto and R. Coelho, editors, *Fuzzy Information Processing. NAFIPS 2018. Communications in Computer and Information Science*, volume 831, pages 290–301, Cham, 2018. Springer International Publishing. Citado 7 vezes nas páginas 14, 16, 57, 68, 76, 77 e 78.
- [68] A. C. Souza and M. E. Valle. Generalized Exponential Bidirectional Fuzzy Associative Memory with Fuzzy Cardinality-Based Similarity Measures Applied to Face Recognition. *Trends in Applied and Computational Mathematics*, 19(2), 2018. Citado 5 vezes nas páginas 16, 68, 81, 82 e 84.
- [69] A. C. Souza and M. E. Valle. Memória associativa bidirecional exponencial fuzzy generalizada com medida de similaridade estrutural aplicada a um problema de reconhecimento de faces. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 6(1), 2018. Citado 3 vezes nas páginas 16, 68 e 86.
- [70] A. C. Souza, M. E. Valle, and P. Sussner. Generalized Recurrent Exponential Fuzzy Associative Memories Based on Similarity Measures. In *Proceedings of the 16th World Congress of the International Fuzzy Systems Association (IFSA) and the 9th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT)*,

- volume 1, pages 455–462. Atlantis Press, 2015. DOI: 10.2991/ifsa-eusflat-15.2015.66. Citado 7 vezes nas páginas 13, 17, 28, 39, 42, 52 e 98.
- [71] P. Sussner, R. Miyasaki, and M. E. Valle. An Introduction to Parameterized IFAM Models with Applications in Prediction. In *Proceedings of the 2009 IFSA World Congress and 2009 EUSFLAT Conference*, pages 247–252, Lisbon, Portugal, July 2009. Citado na página 12.
- [72] P. Sussner and M. E. Valle. Grayscale Morphological Associative Memories. *IEEE Transactions on Neural Networks*, 17(3):559–570, May 2006. Citado na página 12.
- [73] P. Sussner and M. E. Valle. Recall of Patterns Using Morphological and Certain Fuzzy Morphological Associative Memories. In *Proceedings of the IEEE World Conference on Computational Intelligence 2006*, pages 209–216, Vancouver, Canada, 2006. Citado na página 12.
- [74] P. Sussner and M. E. Valle. Fuzzy Associative Memories and Their Relationship to Mathematical Morphology. In W. Pedrycz, A. Skowron, and V. Kreinovich, editors, *Handbook of Granular Computing*, chapter 33, pages 733–754. John Wiley and Sons, Inc., New York, 2008. Citado 2 vezes nas páginas 12 e 28.
- [75] M. E. Valle. Fuzzy Exponential Recurrent Neural Networks for Gray-scale Image Retrieval. In *Proceedings of the Third Brazilian Conference on Fuzzy Systems (III CBSF)*, João Pessoa, Brazil, August 2014. Citado 5 vezes nas páginas 12, 17, 28, 37 e 98.
- [76] M. E. Valle and A. C. Souza. Pattern classification using generalized recurrent exponential fuzzy associative memories. In A. G. H. George A. Papakostas and V. G. Kaburlasos, editors, *Handbook of Fuzzy Sets Comparison Theory, Algorithms and Applications*, volume 6, chapter 4, pages 79–102. Science Gate Publishing, 2016. Citado 5 vezes nas páginas 16, 39, 68, 71 e 72.
- [77] R. Wang, Y. Li, M. W. Mahoney, and E. Darve. Structured block basis factorization for scalable kernel matrix evaluation. 2015. arXiv:1505.00398v1. Citado na página 66.
- [78] Z. Wang and A. C. Bovik. Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures. *IEEE Signal Processing Magazine*, 26(1):98–117, Jan. 2009. Citado na página 27.
- [79] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. DOI:10.1109/TIP.2003.819861. Citado 6 vezes nas páginas 13, 15, 17, 26, 27 e 85.

- [80] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210–227, Feb 2009. Citado 5 vezes nas páginas 15, 68, 69, 90 e 93.
- [81] L. Xuecheng. Entropy, distance measure and similarity measure of fuzzy sets and their relations. *Fuzzy Sets and Systems*, 52(3):305–318, 1992. DOI:10.1016/0165-0114(92)90239-Z. Citado 6 vezes nas páginas 13, 15, 17, 22, 23 e 25.
- [82] X. Yin and J. Han. CPAR: Classification based on predictive association rules. In *Proceedings of the 2003 SIAM International Conference on Data Mining*, pages 331–335. SIAM, 2003. Citado na página 77.
- [83] L. A. Zadeh. Fuzzy Sets. *Information and Control*, 8(3):338–353, 1965. DOI:10.1016/S0019-9958(65)90241-X. Citado na página 19.
- [84] B.-L. Zhang, H. Zhang, and S. S. Ge. Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory. *IEEE Transactions on Neural Networks*, 15(1):166–177, Jan. 2004. Citado 8 vezes nas páginas 12, 17, 28, 34, 36, 37, 90 e 98.
- [85] D. Zhang and W. Zuo. Computational Intelligence-Based Biometric Technologies. *IEEE Computational Intelligence Magazine*, 2(2):26–36, May 2007. Citado na página 12.
- [86] H. Zhang, W. Huang, Z. Huang, and B. Zhang. A Kernel Autoassociator Approach to Pattern Classification. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 35(3):593–606, June 2005. Citado 4 vezes nas páginas 14, 36, 37 e 51.
- [87] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *2011 International Conference on Computer Vision*, pages 471–478, Nov 2011. Citado 2 vezes nas páginas 90 e 93.
- [88] R. Zwick, E. Caristein, and D. V. Budesu. Measures of Similarity Among Fuzzy Concepts: A Comparative Analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, Apr. 1987. Citado na página 25.